AUGUST 1, 1969

COORDINATED SCIENCE LABORATORY

PROGRESS REPORT FOR SEPTEMBER 1968 THROUGH JUNE 1969

UNIVERSITY OF ILLINOIS – URBANA, ILLINOIS

The research reported in this document was made possible through support extended the Coordinated Science Laboratory, University of Illinois, by the Joint Services Electronics Program (U.S. Army Electronics Laboratories and U.S. Army Research Office, Office of Naval Research and the Air Force Office of Scientific Research) under Contract Number DAAB-07-67-C-0199.

Portions of this work were also supported by:

National Aeronautics and Space Administration

Grant NsG 228-62, Grant NAS 8-21442

Jet Propulsion Lab (NASA)

Contract 952383

National Science Foundation

Grant NSF GK-1663, Grant NSF GK-2339, Grant NSF GK-3893

Advanced Research Projects Agency

Through U. S. Army Contract DAAK-02-67-C-0546

Air Force Office of Scientific Research

Grant AFOSR 68-1579, Grant AFOSR 69-1671

Air Force Systems Command

F33615-69-C-1251

Office of Naval Research

Contract N00014-67-A-0305-0001

Naval Air Development Center

N2269-69-C-0013

U. S. Office of Education

Contract OE C-1-7-071213-4557

American Chemical Society

Contract PRF 3668-A3,5

Ford Motor Company

Owens-Illinois, Incorporated

Bi-University Institutional Liaison for Development

Physical Electronics Affiliates Program
College of Engineering, University of Illinois

University of Illinois

as acknowledged in footnotes in the text

COORDINATED SCIENCE LABORATORY
SUMMARY OF
PROGRESS REPORT FOR SEPTEMBER 1968 THROUGH JUNE 1969

1. Surface Physics

During the past year, three experiments have been pursued to the stage of the preparation of Ph.D. theses. These are electron-impact desorption of ions and neutrals, energy-angular distribution of secondary electrons, and high-resolution electron spectroscopy of solid surfaces. The thesis for the first of these has been completed and has been presented as a CSL report. The second two are in preparation. Brief summaries of the work covered in these theses are given in the present progress report. A new activity in surface research has been started in the laboratory under the direction of Professor Gert Ehrlich. This work emphasizes the study of the chemistry of the gas-solid interface. A brief description of the types of experiments which will be pursued is given.

2. Applied Physics

Resistivity and mobility data as a function of temperature, and photoluminescent spectra are given for vacuum-deposited CdS films about one micron thick. Measurements on metal-insulator-CdS structures show light emission at voltages as low as one volt, making a tunnel injection mechanism doubtful. Parameters are discussed for experimental measurements of barrier heights in tunneling experiments. The behavior of an omegatron total-partial pressure gauge is discussed, as well as the effect of small magnetic fields on typical ionization gauges. The scanning electron microscope has been used to examine surfaces prepared by electrical-discharge machining. The results have led to cutting conditions improving the rate, and reducing surface damage.

3. Plasma Physics

The effect of collisions on the instability of cold and warm electron beams in a plasma has been investigated experimentally. It is found that the instability is effectively suppressed for velocity spreads much smaller than previously assumed. Computer calculations are presented that support this finding. The spatial-temporal development of the instability in the presence of boundaries is discussed theoretically.

SUMMARY

4. Rarefied-Gas Dynamics

More accurate Monte-Carlo solutions of the Boltzmann equation have been obtained for the pseudoshock and the heat-transfer problems by using the new techniques developed. These solutions and the solution of the shock wave problem enable us to make detailed studies of the nonequilibrium gas behavior in rarefied-gas flow problems. These studies include the investigation of the effect of the gas-surface interaction on the gas behavior near a solid surface. Because of the improvement of the accuracy of evaluating the Boltzmann collision integrals, we are able to extend our study to near-equilibrium conditions. The second-order effect of the Chapman-Enskog's distribution function on the collision integral is being studied. We have conducted more error studies of our Monte-Carlo method to establish a firm basis for tackling more complex problems. We have completed the initial phase of our plans to extend the application of our Monte-Carlo method to more complex problems by using a large-scale parallel-operation computer.

5. High-Voltage Breakdown

The nature of the cathode protrusions responsible for initiation of electrical breakdown in ultrahigh vacuum has been studied by means of the scanning electron microscope. These protrusions appear to be fingers of metal extending from the edges or interiors of small craters. The craters are produced when clumps of charged material, as yet unidentified, impinge on the surface of the cathode (or the anode), melting the local region. Using drastic cleaning procedures, the clumps, under laboratory conditions, appear to have been eliminated, thus eliminating the protrusions and permitting the electrode voltage to be raised one to two orders of magnitude higher, without field emission or electrical breakdown, than was previously practical. In several cases, 60 kV has been sustained between tungsten electrodes spaced 0.04mm apart, indicating a field of about 15×10^8 V/m at the electrode surface.

6. Space and Atmospheric Sciences

A statement of the status of the feasibility study for the orbiting-gyro relativity satellite is given. The program and objectives of the study of another orbiting-gyro experiment designed to measure gas density and gas-surface interaction parameters are given. The results of a preliminary study of a composite-

shaped satellite are given which illustrate the feasibility of the proposed atmospheric experiment. The resulting equations of a more complete aerodynamic analysis of spinning disks, cones, and cylinders at arbitrary angles of attack and the evaluation of these results using numerical techniques are discussed. The feasibility of measuring the rates of adsorption and desorption of neutral molecules on the satellite surface is evaluated as a possible secondary objective in the proposed atmospheric experiment. Molecular vibrations of cyclododeca-sulfur, S_{12}, were interferometrically measured in the far infrared using a polarized source.

7. Semiconductor Physics

Defects and impurities determine many of the properties of semiconductors. In many cases, radiation offers a convenient means of introducing defects. Results are presented of the influence of irradiation upon the recombination luninescence of silicon, the luminescent spectra of CdSe when excited into laser action, and the electrical properties of p-i-n diodes. The tunneling between metal-semiconductor junctions is investigated as a function of the impurity level in the semiconductor.

8. Computer

Arithmetic and code-checking routines have been added to the ILLAR system. The FORTRAN compiler has been modified to make subprograms recursive (i.e., they can call themselves). The CSLx programming package has been completed, and the Mitchell 16-mm movie production facility has been installed. The display system has been modified to provide 256 intensity levels, and a programming system that uses this feature has been developed to produce prospective drawings in which hidden planes are eliminated. Examples are shown.

9. Plasma Display

The appearance of a 128X128 Plasma Display Panel connected to a digital computer is shown. Studies of two-on-state operation, applications to memory, circuits, field calculations, stability, and color in the Plasma Display Panel are discussed.

SUMMARY

10. Urban Development

An in-depth study of the factors causing the growth and decay of Kankakee, Illinois has been started. During this first phase of the study, data have been collected and prepared for computer display on the location, price, and date of real estate transactions throughout the Kankakee metropolitan area. Plans for the analysis of these data are described.

11. Information Science

Progress in the area of information-processing techniques and systems is reported. Topics discussed include methods of achieving systems reliability through coding, design techniques of digital systems, estimation and filtering, source coding and time-continuous channels, information retrieval and communications.

12. Networks

A computer optimization method is reported for the broad-band matching of complex generator and load impedances to a lumped-distributed network. The network may contain active elements such as transistors. The study of system diagnosis from the topological point of view is in progress. In addition some results from the study of the characterization of biplanar graphs are given.

13. Control Systems

Intensive investigations of computationally feasible methods of design of large-scale systems is underway. Furthermore, additional results concerning minimax algorithms and feedback strategies for differential games have been obtained. Finally, suboptimal filtering, sensitivity of distributed parameter systems, adaptive systems, and applications of optimal control theory to traffic control and rolling-mill design are being investigated.

14. Switching Systems

The investigation of fault diagnosis methods for digital systems has been continued. Results have been obtained on the improvement of diagnosability by test-point insertion, on the detection of multiple faults by single fault tests, and on the generation of diagnostic tests for networks of modules. Work continues on methods of synthesizing sequential networks which are optimal from

the diagnosis viewpoint and on fault-diagnostic properties of regular arrays.

15. Cogntitve Memory

Fundamental properties of perceptive cognition are investigated. Semantic techniques for parsing and compiling are developed. Architecture of an associative processor is presented. Other progress relating to cognitive structures are also reported.

16. Seismic Vehicle Detection

A telemetered engine clock signal is being used in an attempt to determine if vehicular seismic signals contain well-defined frequency components related to engine speed. Preliminary results of an attempt to determine the weight of passing vehicles from an analysis of their seismic signals is described.

17. Engineering Aids for the Handicapped

A preliminary design is described for a highway vehicle to be operated by a severely disabled person. A vehicle is being donated, and a proposal has been written for further study and the modification of the donated vehicle. A technique which may enable blind to read newsprint is described.

18. Photo-Chromic Memory for Computer Applications

A proposed mechanism for using M_A color centers in a write, erase, and nondestructive read-out memory system is described.

Beginning with this issue of the Progress Report, an effort is being made to report all physical quantities in the System of International Units (SI). The basic SI units are

length	meter	m
mass	kilogram	kg
time	second	s
electric current	ampere	A
absolute temperature	kelvin	K
luminous intensity	candela	cd.

Certain purely numeric units are also used, such as,

plane angle	radian	rad
solid angle	stereradian	sr
logarithmic ratio	bell	B,

while the angular degree and its base-60 submultiples sometimes appear in appropriate contexts. Similarly, the usual multiples of the second, e.g., minute, hour, day, etc., are used as appropriate.

Derived SI units with special names are

energy	joule	J	$kg\ m^2 s^{-2}$
force	newton	N	Jm^{-1}
power	watt	W	Js^{-1}
electric charge	coulomb	C	As
electric potential difference	volt	V	$JA^{-1}s^{-1}$
electric resistance	ohm	Ω	VA^{-1}
electric capacitance	farad	F	AsV^{-1}
magnetic flux	weber	Wb	Vs
inductance	henry	H	VsA^{-1}
magnetic flux density	tesla	T	Vsm^{-2}
luminous flux	lumen	lm	$cd\ sr$
illumination	lux	lx	$cd\ sr\ m^{-2}$
frequency	hertz	Hz	cycle per second,

and these are also used, together with the customary celsius (or centigrade) scale of temperatures. Certain non-SI units are used, such as the metric ton (t) and the electron volt (eV), for example. The usual scale of numerical prefixes is used, but with deci- (d), centi- (c), deka- (da), and hecto- (h), being avoided (exception: dB). A name for the unit of electrical conductance has been omitted from the above list, since there is no international symbol other than A/V.

Among the many non-SI units which are being avoided are these:

length	angstrom	0.1 nm
	mil	25.4 μm
	centimeter	10 mm
	inch	25.4 mm
	foot	0.3048 m
	mile	1.60934 km
	nautical mile	1.852 km
area	square centimeter	10^{-4} m^2
volume	cubic centimeter	10^{-6} m^3
mass	pound	0.453 592 37 kg
force	dyne	10^{-5} N
pressure	torr	133.322 Nm^{-2}
energy	erg	10^{-7} J
	calorie (IT)	4.1868 J
	BTU	1055.06 J
power	horsepower	745.700 W.

In some of the illustrative material prepared early in this reporting period, however, the legends occasionally indicate the use of some of these noninternational units.

No effort has been made, for quantities having primarily an economic or legal significance, to report them in the System of International Units.

D. H. Cooper
Progress Report Editor

COORDINATED SCIENCE LABORATORY PERSONNEL

Faculty, Research Associates, and Research Engineers

Compton, W. D., Director
Alpert, D.
Anner, G. E.
Ash, R. B.
Bitzer, D. L.
Bohmer, H.
Bouknight, W. J.
Bullen, A. G. R.
Carroll, D. E.
Carss, B. W.
Chien, R. T.
Coates, W. J., Mr.
Cooper, D. H.
Cordaro, J. T., Jr.
Cruz, J. B., Jr.
Culton, J. W.
Ettinger, S. Y.
Fenves, S. J.
Gooch, J. D.
Haddad, A. H.
Hohn, F. E.
Jackson, E. A.
Kadanoff, L. P.
Kelley, K. C.
Kirkwood, B. D.
Knoebel, H. W.
Kokotovic, P.
Krone, H. V.
Lee, D. A.
Lyman, E. M.
Lyman, E. R.
Mayeda, W.
Medanic, J.
Metze, G.
Peacock, R. N.
Peisl, H.
Perkins, W. R.
Preparata, F. P.
Propst, F. M.
Prothe, W. C.
Raether, M.
Ray, S. R.
Resch, R. D.
Scott, E. J.
Skaperdas, D.
Slottow, H. G.
Steinrisser, F.
Stifle, J. E.
Stoneham, A. M.
Trick, T. N.
Trogdon, R.
Tulumello, A.
Utkin, V.
Vlach, J.
Von Foerster, H.
Voss, J. R.
Voth, B. W.
Warwick, J. W.
Wax, N.
Yen, S. M.

Assistant to the Director

Research Assistants

Akin, J. G.
Allen, R. C.
Anderson, T. E.
Arnold, P. J.
Arora, B. M.
Ayrault, G.
Bahl, L. R.
Birtcher, R. C.
Biss, K. O.
Bleha, W. P., Jr.
Bollinger, L. D.
Brown, W. F.
Burkstrand, J. M.
Burt, J. V.
Chambers, R. S.
Chang, J. S.
Chen, C. I.
Chow, J. C.
Chow, W.
Chu, W. H.
Cohen, A. Z.
Cooper, T. L.
Cullen, D. E.
Cummings, J. W.
D'Ans, G.
Davis, J. M.
Depp, S. W.
Duran, P. S.
Edwards, D., Jr.
Evans, K. E., Jr.
Fenneman, D. B.
Foster, C. A.
Gardner, M. F.
Hartman, W. H.
Hartmann, C. R.
Hayes, J. M.
Hecht, R. W.
Heller, J. E.
Hinata, S.
Holden, R. A.
Hong, S. J.
Ito, M.
Jamshidi, M.
Jimenez, R. C.
Johnson, E. S.
Johnson, H. R.
Johnson, R. L.
Jones, C. E.
Karr, G. R.
Kovener, G. S.
Kraybill, D. M.
Lipovski, G. J.
Livingston, D. C.
Lombardi, D. J.
Mark, E. A.
Martin, R. J.
Metze, V. C.
Miller, M. K.
Moore, J. K.
Moy, A. B.
Ng, S. W.
Nicholas, M. J.
Nishijima, M.
Pavek, R. E.
Petty, W. D.
Polizzotti, R. S.
Powell, T.
Reilly, B.
Rice, H. M.
Robinson, J. E.
Rowe, R. G.
Sakurai, J.
Schein, L. B.
Schermerhorn, J. D.
Schertz, D.
Schultz, J. A.
Seth, S. C.
Stahl, F. A.
Stewart, C. N.
Stowe, D. W.
Stupp, G. B.
Tao, B.
Taylor, S. M.
Toida, S.
Tzeng, K. K.
Wallat, R. J.
Wendelken, J. F.
Weston, P. E.
Wilkins, M. G.
Withrow, S. P.
Wojcik, A. S.
Woo, T.
Wrigley, J. D., Jr.
Wu, V.
Yamada, M.

PERSONNEL

Fellows

Cullum, R. D.
Finkelstein, H. A.
Gaddess, T.
Miller, D. L.
Mokari-Bolhassan, M. E.
Paige, M. R.

Administrative Secretary

Schmidt, R. F.

Typists and Stenographers

Anson, J.
Champagne, B.
Hanoka, N.
Kallembach, S.
Kempton, B.
Lahey, M.
Lane, R.
Siler, J.
Stillings, N.

Chief Clerk

Drews, C.

Duplicating Machine Operator

Potter, R.

Chief Engineering Draftsman

MacFarlane, R.

Draftsman

Kimbrell, R.

Research Lab Shop Supervisor

Bandy, L.

Research Engineering Asst.

Burr, J. G.

Laboratory Mechanics

Bales, R.
Beaulin, W.
Bouck, G.
Fults, R.
Merritt, K.

Electronics Engr. Assts.

Carter, E.
Gardner, O. E.
Hedges, L.
Neff, E.
Vassos, N.

Electronics Technicians

Casale, T.
Coad, D.
Crawford, G.
Deschene, D.
Holy, F.
Johnson, M.
Knoke, J.
Merrifield, F.
Moule, G.
Roberts, G.
Schmidt, W.
Streff, L.
Susedik, A.
Turpin, F.

Physical Sci. Staff Asst.

Thrasher, W.

Stores Supervisor

Lofton, C.

Storekeepers

Jordan, F.
McElwain, W.

Glassblower

Lawrence, W.

Photographer

Gladin, R. T.

Student Assistants

Armstrong, M.
Arnold, C.
Belssner, E.
Bonacci, J.
Burkholder, R.
Burnitz, R.
Dehn, M.
Durbin, E.
Feiza, A.
Flood, J. R.
Gates, M.
Goodman, G.
Hansen, J.
Hanson, D.
Jones, M.
Kubala, R.
Loew, F.
McGuire, E.
Payne, B.
Plunk, M.
Reynolds, P.
Scott, R.
Smith, M.
Szymanski, P.
Tucker, P.
Ward, W.
Yeaney, D.
Yonke, M.
Zbylut, R.

1. Journal Articles Published or Accepted

L. R. Bahl and R. T. Chien, "On Gilbert Burst-Error-Correcting Codes," Trans. on Inf. Theory IT-15, 431 (May 1969).

H. Böhmer, J. Chang, and M. Raether, "Collisional Damping of the High-Frequency Instability in a Beam-Plasma System", Plasma Physics (in press) (1969).

H. Böhmer and M. Raether, "Observation of Surface Modes Excited by an Electron Beam in a Plasma," Physics Letters (in press) (1969).

J. J. Bourquin and T. N. Trick, "Stability of a Class of Lumped-Distributed Systems," Journal of the Franklin Institute 287, (in press) (May 1969).

J. Chang, H. Böhmer, and M. Raether, "Generation of Smeared-out Electron Beams for Experiments on Weak-Plasma Turbulence," Rev. Sci. Instr. 39, 1873 (Dec., 1968).

R. T. Chien, "Burst-Error Correction with High-Speed Decoding," IEEE Trans. on Inf. Theory 15, 109 (Jan., 1969).

R. T. Chien, L. R. Bahl, and D. T. Tang, "Codes for Correcting Two Erasure Bursts," IEEE Trans. on Inf. Theory 15, 186 (Jan., 1969).

R. T. Chien, B. Cunningham, and I. Oldham, "Hybrid Methods for Finding Roots of a Polynomial--with Application to BCH Decoding," IEEE Trans. on Inf. Theory 15, 329 (March 1969).

W. D. Compton and R. J. Spry, "Recombination Luminescence in Irradiated Silicon," Phys. Rev. 175, Number 3 (1968).

D. H. Cooper, "Hertzian Contact-Stress Deformation Coefficients," J. Applied Mech., (in press) (1969).

D. E. Cullen and W. D. Compton, "Tunneling Spectroscopy in Highly P-Type Silicon," Bull. Amer. Phys. Soc. 14, 414 (March 1969)

E. S. Davidson and G. Metze, "Comments on 'An Algorithm for Synthesis of Multiple-Output Combinational Logic'," IEEE Trans. on Computers C-17, 1091 (Nov., 1968).

A. H. Haddad, "Nonlinear Prediction of a Class of Random Processes," IEEE Trans. on Information Theory IT-14, 664 (Sept., 1968).

A. H. Haddad, with J. B. Thomas and D. S. Kuykendall, "Series Approximations for Optimum ZNL Filters," IEEE Trans. on Circuit Theory CT-15, 201 (September, 1968).

P. Kokotović and J. B. Cruz, Jr., "An Approximation Theorem for Linear Optimal Regulators," Journ. of Mathematical Analysis and Applications 27, to appear (July 1969).

P. Kokotović, W. R. Perkins, J. B. Cruz, Jr., and G. D'Ans, "ε-Coupling Method for Near-Optimum Design of Large-Scale Linear Systems," Proc. of the IEE (London) 116, 889 (May 1969).

V. Lum and R. T. Chien, "On the Minimum Distance of Bose-Chaudhuri-Hocquenghem Codes," SIAM J. Applied Math. 16, 1325 (Nov. 1968).

E. A. Mechtly, M. M. Rao, D. O. Skaperdas, and L. G. Smith, "Latitude Variation of the Lower Ionosphere," Radio Science 4 (June, 1969).

I. Oldham, R. T. Chien, and D. T. Tang, "Error Correction in a Photo-Digital Mass-Storage System," IBM J. Research and Development 12, 422 (Nov., 1968).

T. J. Powell, "A Procedure for Selecting Diagnostic Tests," IEEE Trans. on Computers C-18, 168 (Feb., 1969).

F. P. Preparata, "A Class of Optimum Nonlinear Double Error-Correcting Codes," Information and Control 13, 378 (Oct., 1968).

F. P. Preparata, "An Estimate of the Length of Diagnostic Tests," IEEE Transactions on Reliability (in press).

P. Sannuti and P. Kokotović, "Near-Optimum Design of Linear Systems by a Singular Perturbation Method," IEEE Trans. on Automatic Control AC-14, 15 (February, 1969).

D. Salmon and P. Kokotović, "Design of Feedback Controllers for Nonlinear Plants," IEEE Trans. on Automatic Control AC-14, (June, 1969).

F. Steinrisser, L. C. Davis, and C. B. Duke, "Electron and Phonon Tunneling Spectroscopy in Metal-Germanium Contracts," Phys. Rev. 176, Number 3 (1968).

B. G. Streetman, N. Holonyak, Jr., H. V. Krone, and W. D. Compton, "Current Oscillations in Si p-i-n Devices After Irradiation with One-MeV Electron," Appl. Phys. Letters 14, 63 (1969).

D. T. Tang and R. T. Chien, "Error Control with Coding," IBM Systems Journal 8, 48 (January, 1969).

T. N. Trick, "Asymptotic Stability of a Class of Networks Containing a Nonlinear Time-Varying Capacitor and Periodic Inputs," IEEE Trans. on Circuit Theory CT-16, 217 (May 1969).

T. N. Trick and D. R. Anderson, "Stability of Nonlinear Networks," IEEE Trans. on Circuit Theory CT-16, (in press) (August 1969).

T. N. Trick and J. J. Bourquin, "Transient Analysis of Distributed Systems," Proc. IEEE 57 (in press) (June 1969).

N. Wax, "A Resolution of a Circuit Paradox," Proc. IEEE 56, 2073 (Nov., 1968).

N. Wax, "A Spate of 'Paradoxes,'" Proc. IEEE 57, 711 (April, 1969).

D. F. Wilkie and W. R. Perkins, "Generation of Sensitivity Functions for Linear Systems Using Low Order Models," IEEE Trans. on Automatic Control, (in press)(April, 1969).

E. L. Wolf and W. D. Compton, "A Practical Evaporator Fixture for Vacuum Cleavage of Semiconductor Crystals," Rev. Sci. Instr., (in press).

S. M. Yen, "Numerical Methods for Solving Rarefied Gas Flow Problems," Applied Mechanics Reviews 22, 557 (June, 1969).

2. Meeting Papers

H. Böhmer, J. Chang, and M. Raether, "Collisional Quenching of the Beam-Plasma Instability for Electron Beams with Large Velocity Spread," Symposium on Beam-Plasma Interactions, Utrecht, June 1969.

H. Böhmer and M. Raether, "Observation of Surface Modes Excited by an Electron Beam in a Plasma," Meeting Div. of Plasma Physics, APS, Miami, November 1968.

R. T. Chien and F. P. Preparata, "Search Strategy and File Organization in Computerized Information-Retrieval Systems with Mass Memory," Proc. of the FID/IFIP Conf. 1967 on Mechanized Information Storage Retrieval and Dissemination, pp. 108-122, July 1968.

R. T. Chien and K. K. Tzeng, "On Cyclic Codes for Correcting Solid-Burst Errors, Proc. Third Ann. Princeton Conf. on Information Sciences and Systems, March 1969.

E. S. Davidson and G. Metze, "Module Complexity and NAND Network Design Algorithms," Proc. 6th Annual Allerton Conf. on Circuit and Systems Theory, Oct., 1968, pp. 538-548.

PUBLICATIONS

S. Y. Ettinger and E. M. Lyman, "Prebreakdown Processes between Plane Electrodes in Clean Vacuum Systems," Proc. Third Int. Symposium on Discharges and Electrical Insulation in Vacuum, pp. 128-133, Paris, Sept., 1968.

A. H. Haddad, with J. V. Bodycomb, "Some Properties of a Predictive Quantizing System," The 3rd Ann. Princeton Conference on Information Sciences and Systems, Marcy 1969.

A. H. Haddad, with P. V. Kokotović and J. B. Cruz, Jr., "Design of Control Systems with Uncertain Parameters," The 3rd Ann. Princeton Conference on Information Sciences and Systems, March 1969.

J. P. Hayes, "An Algorithm for Determining the Equivalence of Regular Expressions," Proc. 6th Annual Allerton Conf. on Circuit and Systems Theory, pp. 732-742, Oct., 1968.

B. L. Hicks and S. M. Yen, "Calculations of the Boltzmann Collision Integrals for Rarefied-Gas-Flow Problems," APS Meeting (Fluid Dynamics Division), Seattle, Washington, Nov., 1968.

B. L. Hicks and S. M. Yen, "Solution of the Nonlinear Boltzmann Equation for Plane Shock Waves," Proceedings of the 6th International Symposium on Rarefied Gas Dynamics, pp. 313-317, 1969.

P. Kokotović and G. D'Ans, "Parameter Imbedding Design of Linear Optimal Regulators," 3rd Princeton Conf. on Information Sciences and Systems, March 1969.

P. V. Kokotović, W. R. Perkins, J. B. Cruz, Jr., and G. D'Ans, "A Decoupling Method for Approximate Solution of Large Scale Riccati Equations," to appear in the Proc. of the Hawaii International Conf. on System Sciences, January 1969.

J. Medanić, "Elimination Algorithm for Computation of the Minimax," Proc. Sixth Annual Allerton Conf. on Circuit and System Theory, pp. 702-710, October 1968.

J. Medanić and C. I. Chen, "Segment Method for the Control of Systems in the Presence of Uncertainty," Joint Automatic Control Conference, Boulder, Colorado, August 1969.

W. Mayeda, "Distinguishability Criteria and Diagnosis in Computers II," Proceedings of the 12th Midwest Symposium on Circuit Theory, University of Texas, April, 1969.

W. Mayeda and C. V. Ramamoorthy, "Distinguishability Criteria in Oriented Graphs and its Application to Computer Diagnosis," 1968 IEEE International Symposium on Circuit Theory, Dec., 1968.

T. J. Powell, "Synthesis Requirements for Fault Detection," Proc. 6th Annual Allerton Conf. on Circuit and Systems Theory, pp. 761-772, Oct., 1968.

F. P. Preparata, "An Estimate of the Length of Diagnostic Tests," Proc. 3rd Princeton Conf. in Information Sciences and Systems, (in press) March 1969.

F. P. Preparata, "The State Coding Problem of Sequential Networks Revisited," (invited paper) Proc. 12th Midwest Symposium on Circuit Theory, Austin, April 1969.

M. Raether, "Problems in Beam-Plasma Interactions," 4th Tokyo Summer Institute, Kyoto, Japan, September 1968.

P. Sannuti and P. Kokotović, "Singular-Perturbation Method for Near-Optimum Design of High Order Nonlinear Systems," Proceedings of the IV Congress of IFAC, Warsaw, Poland, June 1969, Vol. 68, 70-81. Also presented at the Joint Automatic Control Conference, Boulder, Colorado, August 1969.

D. R. Schertz and G. Metze, "On the Indistinguishability of Faults in Digital Systems," Proc. 6th Annual Allerton Conf. on Circuit and Systems Theory, Oct., 1968, pp. 752-760.

T. N. Trick and J. Vlach, "Computer-Aided Design of Distributed Networks with Complex Terminations," Second Biennial Conference on Engineering Applications of Electronic Phenomena-Computerized Electronics, Cornell University, August 26-28, 1969.

D. F. Wilkie and W. R. Perkins, "Design of Model Following Systems Using the Companion Transformation," Proc. of the Fourth Congress of the International Federation of Automatic Control, Warsaw, Poland, 1969. Also presented at Joint Automatic Control Conference, Boulder, Colorado, August 1969.

D. F. Wilkie and W. R. Perkins, "Generation of Sensitivity Functions for Linear Systems Using Low Order Models," Proc. Sixth Annual Allerton Conf. on Circuit and System Theory, pp. 598-614, 1968.

A. Wojcik and G. Metze, "A Reexamination of Ternary Switching Theory," Proc. 6th Annual Allerton Conf. on Circuit and Systems Theory, pp. 711-721, Oct., 1968.

S. M. Yen and H. J. Schmidt, "Monte Carlo Solutions of the Boltzmann Equation for Heat Transfer Problems," Proceedings of the 6th International Symposium on Rarefied Gas Dynamics, pp. 205-213, 1969.

3. Technical Reports

R-385 Effect of Hole Diameter, Electrode Width and Glass Thickness on Properties of the Plasma Display Cell, Walter F. Goede (July, 1968).

R-390 An Associative Processor for Information Retrieval, Andrew Philip Kisylia (August, 1968).

R-391 A Study of a Feedback Time-Sharing System, Edward J. Beach (September, 1968).

R-392 Simulation of a West Process Cement Rotary Kiln, Jai Hari Dalmia (September, 1968).

R-393 Cylinders: A Data Structure Concept Based on Rings, P. Weston and S. M. Taylor (September, 1968).

R-394 Spacecraft Attitude Control Using Feedback Parameter Optimization, George C. Schvmaker (October, 1968).

R-395 An Application of Computer Optimization Techniques to Network Synthesis, T. J. Aprille, Jr. (October, 1968).

R 396 Optimization of A Gyroscopic Satellite General-Relativity Experiment, James L. Meyers, Jr. (September, 1968).

R-397 Correction of Single and Multiple Bursts of Error, Lalit Rai Bahl (October, 1968).

R-398 An Interactive Document Retrieval System, D. E. Carroll, R. T. Chien, K. C. Kelley, F. P. Preparata, P. Reynolds, S. R. Ray and F. A. Stahl (December, 1968).

R-399 Syntactic Analysis for the R2 System, Kenneth Biss (December, 1968).

R-400 A Study of Electroluminescent Cadmium Sulfide Diodes, Jerry Albert Kastning (December, 1968).

R-401 Temperature Dependence of the Restivity and Hall Effect of Thin CDS Films, Ricardo Luis Jimenez (December, 1968).

R-402 Numerical Studies of the Nonlinear Boltzmann Equation; Part I: Summary of Programs and Theory Used in Monte Carlo Solutions of the Boltzmann Equation, Barbara J. Reilly (January, 1969).

R-403 Electrical Properties of Some Vacuum Deposited Insulating Films, Owen S. Reilly (January, 1969).

R-404 On Iterative Decoding of BCH Codes and Decoding Beyond the BCH Bound, Kenneth Kai Ming Tzeng (January, 1969).

R-405 A New Design Technique of Sub-Optimal Nonlinear Automatic Feedback Control Systems Using Sensitivity Functions, Mohammad S. Jamshida (January, 1969).

R-406 A Gradient Algorithm for Minimax Design, James E. Heller (January, 1969).

R-407 ISL--A String Manipulating Language, K. C. Kelley, S. R. Ray, F. A. Stahl (February, 1969).

R-408 Context Modeling in a Cognitive Memory, Daniel Joseph Lombardi (February, 1969).

R-409 Improving the Diagnosability of Modular Combinational Logic by Test Point Insertion, Terry Gene Gaddess (March, 1969).

R-410 Electron Impact Desorption of Ions and Heutrals from Polycrystalline Tungsten, Mitsuaki Mishijima (April, 1969).

R-411 Observation of Surface Modes Excited by an Electron Beam in a Plasma, H. Böhmer and M. Raether (April, 1969).

R-412 Numerical Studies of the Nonlinear Boltzmann Equation; Part II: Studies of New Techniques of Error Reduction, B. L. Hicks, S. M. Yen, and B. S. Reilly (April, 1969).

R-413 A Module Diagnostic Procedure for Combinational Logic, Theo S. Powell (April, 1969).

R-414 Generation of Diagnostic Tests Using Prime Implicants, Michael Reid Paige (May, 1969).

R-415 A Characterization of Biplanar Graphs, Shunichi Toida (May, 1969).

R-416 A Proof of Tutte's Realizability Condition, Wataru Mayeda (May, 1969).

R-417 Some Results in the Theory of Arithmetic Codes, R. T. Chien, S. J. Hone and F. P. Preparata (May, 1969).

R-418 On the Representation of Digital Faults, Donald R. Schertz (May 1969)

R-419 One-Electron and Phonon-Assisted Tunneling in N-Ge Schottky Barriers, L. C. Davis and F. Steinrisser (June, 1969)

R-420 Tunneling Measurement of Electron-Phonon Interactions in Degenerate Semiconductors, C. B. Duke, M. J. Rice, and F. Steinrisser (June, 1969).

Page

F. Propst
G. Ehrlich
G. Ayrault
J. Burkstrand
T. Cooper
D. Edwards
C. Foster
M. Nishijima
R. Polizzotti
R. Rowe
J. Sakuri
C. Stewart, Jr.

1.1. Electron-Impact Desorption from Polycrystalline Tungsten†

As reported previously [1], a cylindrical magnetic spectrometer has been constructed for the study of electron-impact desorption of ions and neutrals from solid surfaces. The apparatus has been applied to the study of polycrystalline tungsten with O_2, CO, CO_2, N_2, H_2O and H_2 adsorbed. The types of measurements which have been made are ion energy distributions, relative cross sections, adsorption and desorption kinetics, and threshold energies.

Some of the results have been reported at the APS Southwestern meeting [2] and the 29th Physical Electronics Conference [3]. These features of the results obtained will be described below. A detailed report of the work is given in a recent thesis [4].

1.1.1. Apparatus

The cylindrical magnetic spectrometer (Fig. 1.1) is essentially a small β-ray spectrometer of the axially symmetric type. The spectrometer, which is housed in an ultrahigh-vacuum system, incorporates an electron gun, a target, an angle-resolving aperture, an energy-resolving aperture, an electrostatic lens system, and a collector (electron multiplier). The electron gun delivering the incident electrons is placed along the axis of the apparatus. Ions produced at the target follow helical trajectories in the axial magnetic field applied by the solenoid. Only those ions in a particular range of angle and energy pass through the apertures. These are focused onto an electron multiplier by an electrostatic lens system. The

†Supported principally by the Joint Services Electronics Program (U.S. Army, U.S. Navy, U.S. Air Force) under contract DAAB-07-67-C-0199 and in part by the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

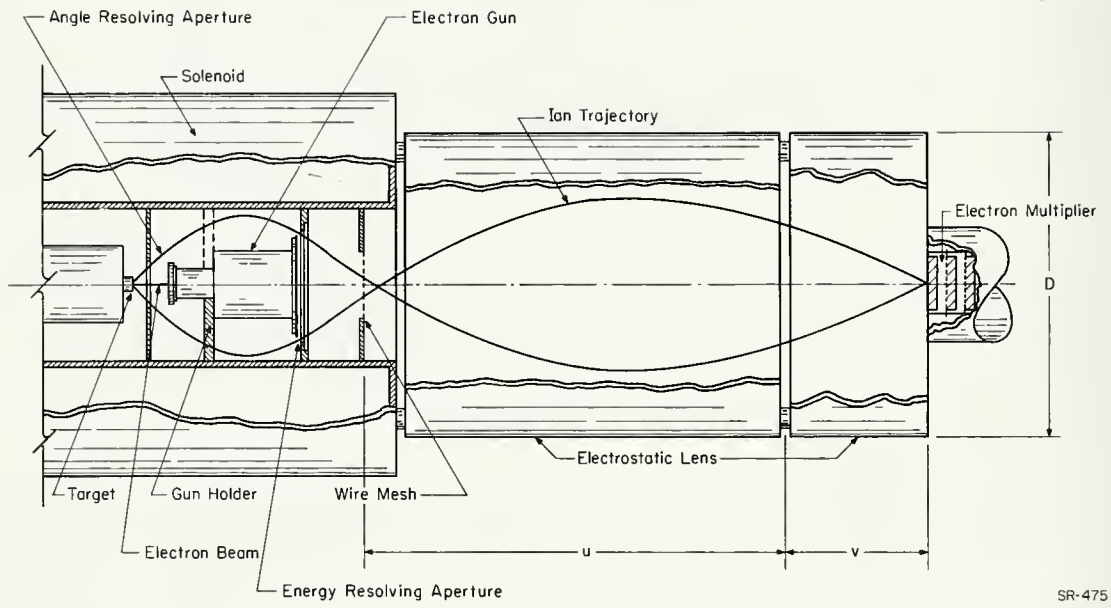


Fig. 1.1. Schematic of the cylindrical magnetic spectrometer. An ion trajectory is also shown.

optimum positions of apertures and the optimum diameter of the source are calculated from the work of DuMond [5].

1.1.2. Ion Species

The technique for the identification of the mass of the emitted ions is illustrated in Fig. 1.2. The energy distributions of ions emitted from the CO/W system are shown for two values of the bias between the target and the analyzer. From the shift of the energy distribution with respect to the change in the bias voltage, one can calculate the mass of the ions. In this case the high energy peak has a mass of 15 and the lower peak a mass of 27.4. These values are determined on the basis of the design parameters of the instrument and have not been corrected by a calibration run. A calibration would give closer values for the masses. The precision is obviously sufficient to identify the two ions in the distribution from CO/W as O^+ and CO^+. The ions observed in the other systems are

O^+ - O_2/W O^+ - CO_2/W N^+ - N_2/W

O^+, CO^+ - CO/W N^+ - N_2O/W

No other ions were observed in any of the systems. The N_2/W system was also studied but no ions were observed. The ionization probability for this system is less than 10^{-9} ions per electron.

1.1.3. Ion-Energy Distributions

The ion-energy distribution for O^+ ions from O_2/W is shown in Fig. 1.3. The distribution is quite symmetrical about the maximum, has a half width of 3 eV, and has an energy of 8.8 eV at the maximum. The energy scale is corrected by 1.7 eV for the contact potential between the target and the analyzer. This correction is determined by studying the elastic reflection of electrons from the target. The contact potential between the target and the cathode is determined by the retarding-potential technique. Figure 1.4 shows several energy distributions of O^+ ions for various primary electron energies. For $V_e \lesssim 30$ eV, the upper end of the distributions are at various energies. In fact, the upper cutoff is proportional to the electron energy as predicted by the simple theory. For $V_e \gtrsim 30$ eV, the distribution simply rises with energy, the shape being almost independent of the energy.

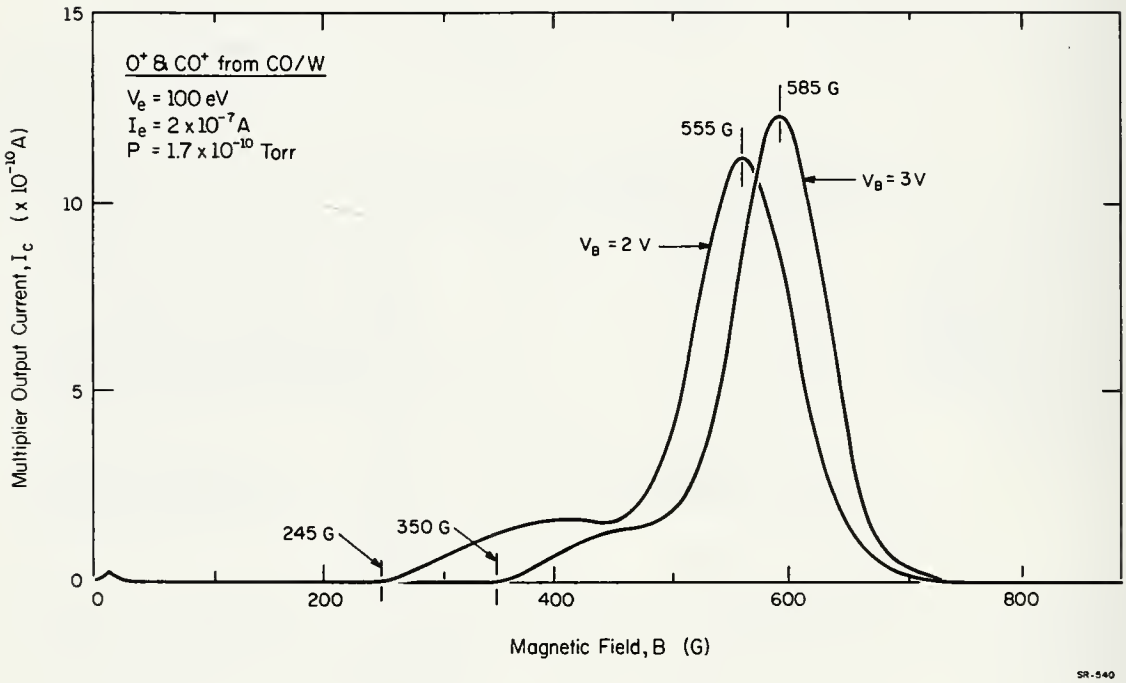


Fig. 1.2. Energy distributions of O^+ and CO^+ ions from CO/W for two values of the target bias. The curves illustrate the method of the mass identification.

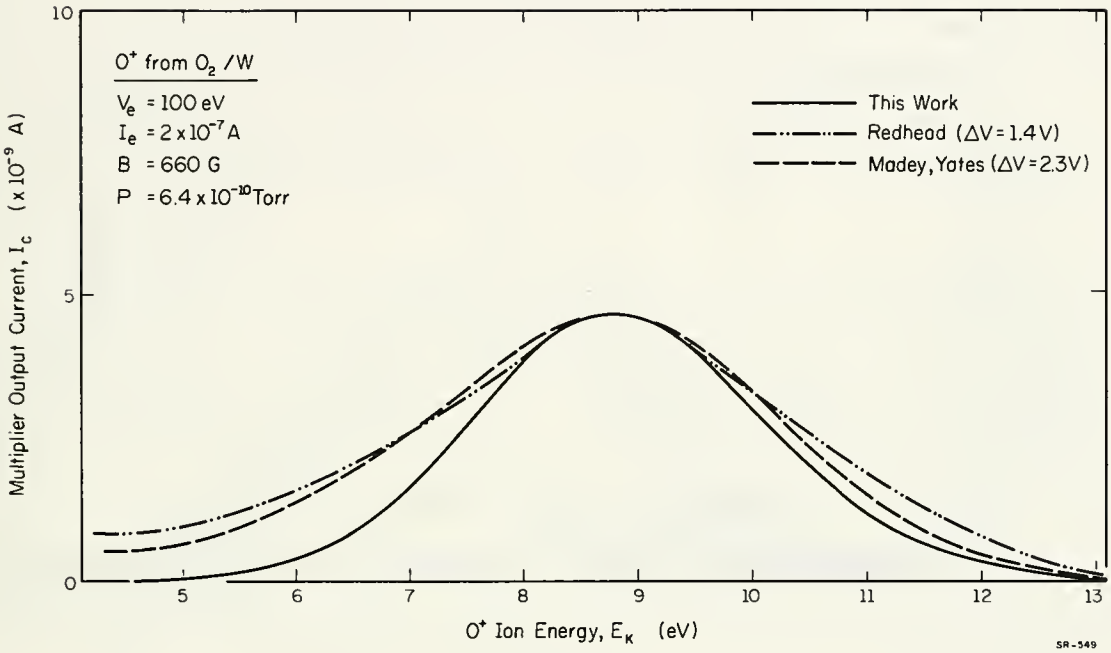


Fig. 1.3. Energy distribution of O^+ ions from tungsten fully covered with O_2. The energy scale is corrected for the contact potential between the target and the analyzer. For comparison, the data of Redhead [6], and Madey and Yates [7], are included. The distributions are normalized and plotted such that the maximum are at the same energy. The energy shift necessary to accomplish this is indicated by ΔV.

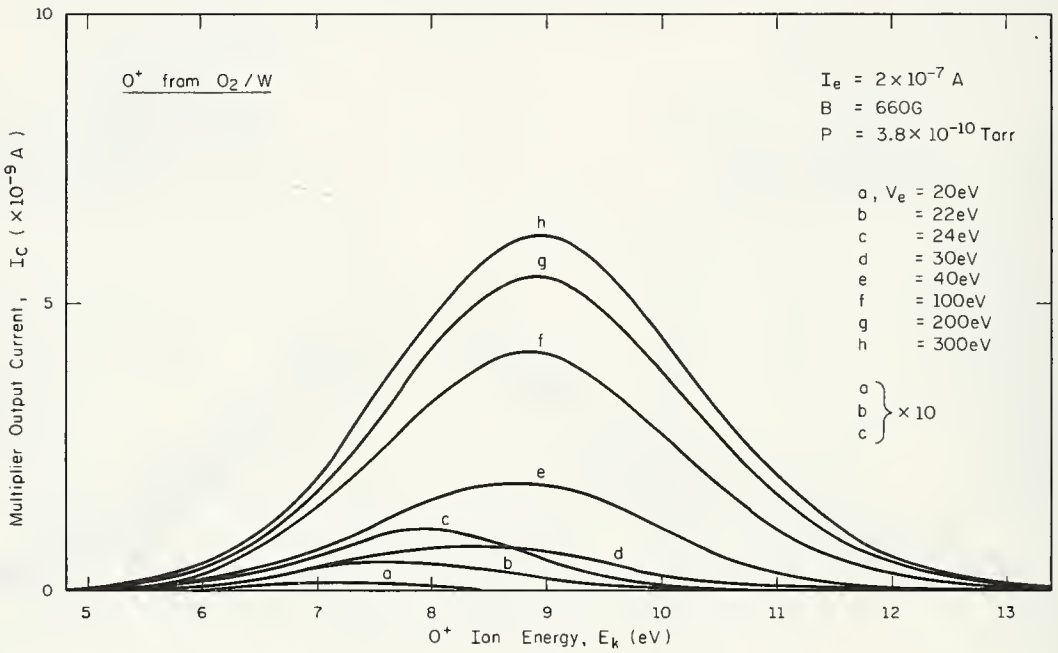


Fig. 1.4. Energy distributions of O^+ ions from O_2/W for various incident electron energies. The electron energy is not corrected for the contact potential between the target and the cathode of the electron gun.

Figures 1.5 and 1.6 show the energy distributions of O^+ and CO^+ ions from CO/W respectively. The shape of the O^+ ion distribution is very similar to the O^+ ion distribution from O_2/W. The maximum is at 7.9 eV. The distribution is symmetrical about this maximum, and the half width is approximately 2.8 eV. The CO^+ ion energy distribution is centered at 1.3 eV and has a half width of 1.6 eV. The energy distributions obtained by other workers are shown on these figures for comparison. It will be noted that in general the major discrepancy between our work and that of others is at lower energies. This is a result of the fact that the fields near the target in the retarding-field analyzers tend to distort the distributions towards lower energies.

The energy distributions of H^+ ions emitted from H_2/W and H_2O/W are shown in Figs. 1.7 and 1.8 respectively. It will be noted that the distributions are at lower energies and not symmetrical as with the oxygen ions observed in other systems. It will also be observed that the shape of the H^+ distributions are quite different in the two cases. The total yield of H^+ ions for H_2O/W is approximately 100 times greater than for H_2/W.

1.1.4. Ion Desorption Cross Sections and Total Desorption Cross Sections

Figure 1.9 shows the ion current for 8.8 eV O^+ as a function of electron energy, V_e, from oxygen covered tungsten. For $V_e \lesssim 100$ eV, the O^+ ion current increases sharply with V_e, and for $V_e \gtrsim 100$ eV, it increases very slowly. The curves obtained by other workers for surface ionization and for gas phase ionization are also shown for comparison. The fact that the measurements of other workers are relatively lower than our results at higher electron energies is again interpreted to be due to the analyzer effect. Figure 1.10 shows the same curves for O^+ ions for CO/W. Again, there is disagreement for probably the same reason. The dependence of the CO^+ ion current from CO/W, H^+ ion current from H_2/W, and H_2O/W on electron energy has also been measured. Except for the H^+ ion current from H_2/W, the dependence is similar to that of O^+ ion current from O_2/W. The H^+ ion current from H_2/W continues to increase as a function of electron energy beyond $V_e \simeq 100$ eV. The H^+ ion ionization probability is estimated to be 7×10^{-9} ions per electron at $V_e = 300$ eV.

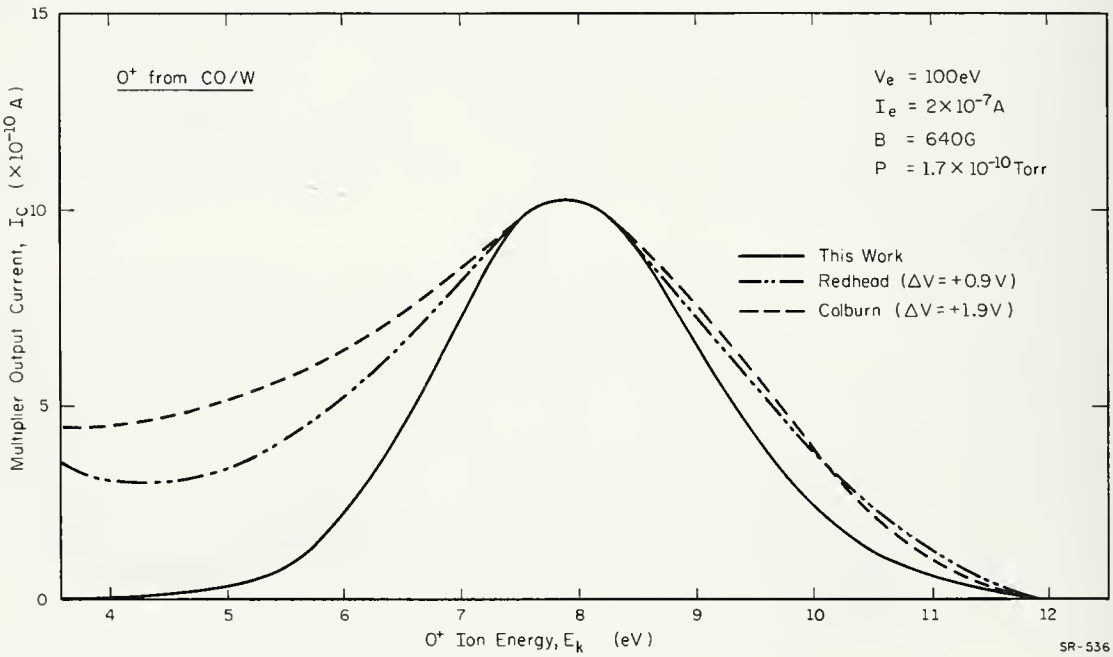


Fig. 1.5. Energy distributions of O^+ ions from tungsten fully covered with CO. The energy scale is set so that the lower edge of the distribution of CO^+ ions is at zero eV. For comparison, the data of Redhead [6], and Coburn [8] are included. The distributions are normalized and plotted such that the maximum are at the same energy. The energy shift necessary to accomplish this is indicated by ΔV.

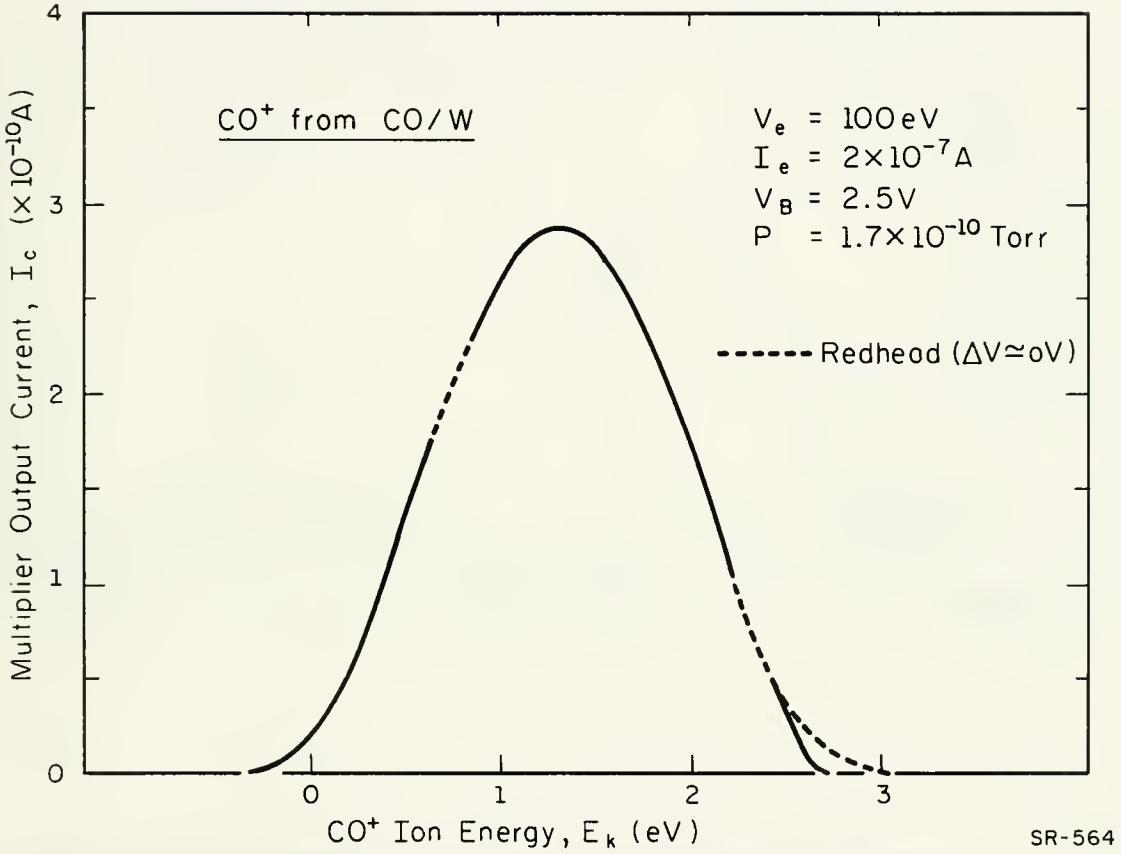


Fig. 1.6. Energy distribution of CO^+ ions from tungsten fully covered with CO. The energy scale is set so that the lower edge of the distribution is at zero eV. For comparison, the data of Redhead [6] are included. The distributions are normalized and plotted such that the maximum are at the same energy. The corresponding energy shift is indicated by ΔV.

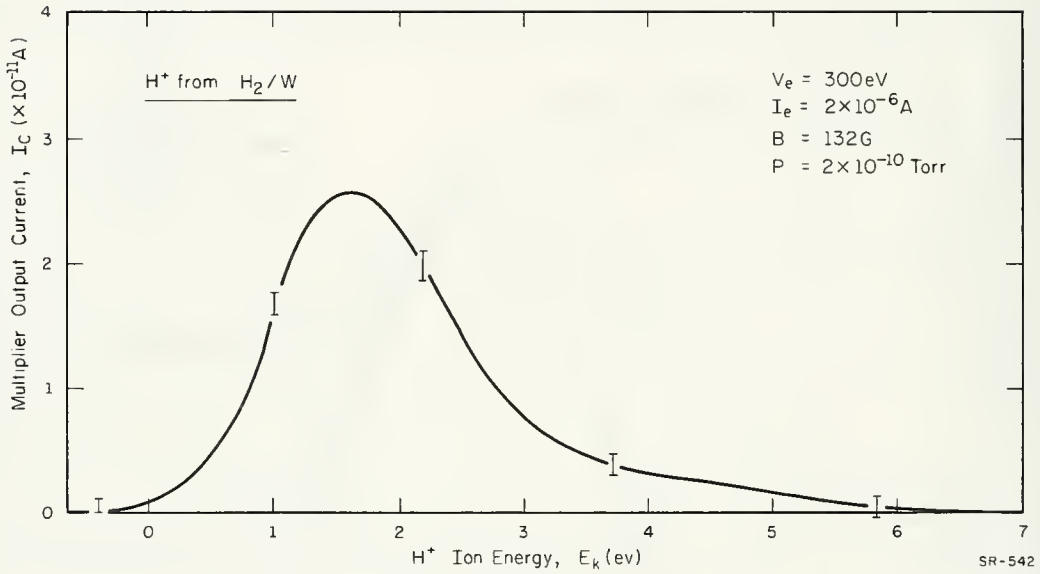


Fig. 1.7. Energy distribution of H^+ ions from tungsten fully covered with H_2. The energy scale is set so that the lower edge of the distribution is at zero eV.

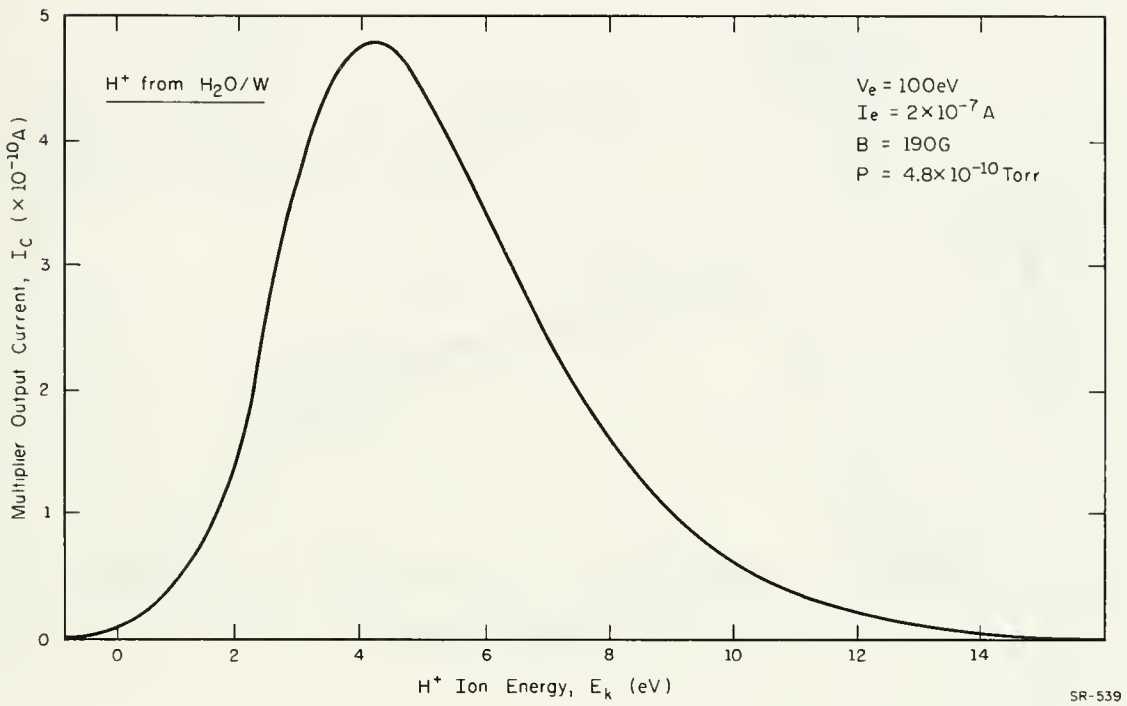


Fig. 1.8. Energy distribution of H^+ ions from tungsten covered with H_2O. The energy scale is set so that the lower edge of the distribution is at zero eV.

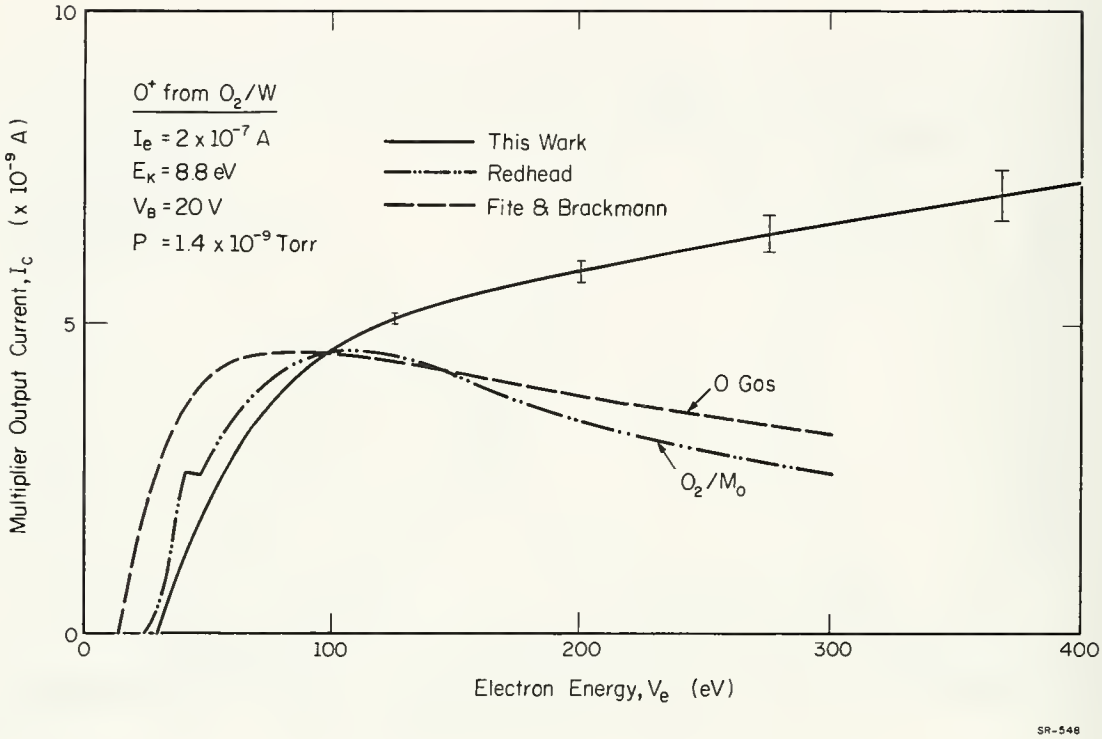


Fig. 1.9. Dependence of the O^+ ion current from O_2/W on the electron energy. Data for O_2/Mo from Redhead [9], and data for free oxygen atoms from Fite and Brachmann [10]. The O^+ ion currents are normalized at V_e = 100 eV. The energy scale for O_2/W is not corrected for the contact potential between the target and the cathode of the electron gun.

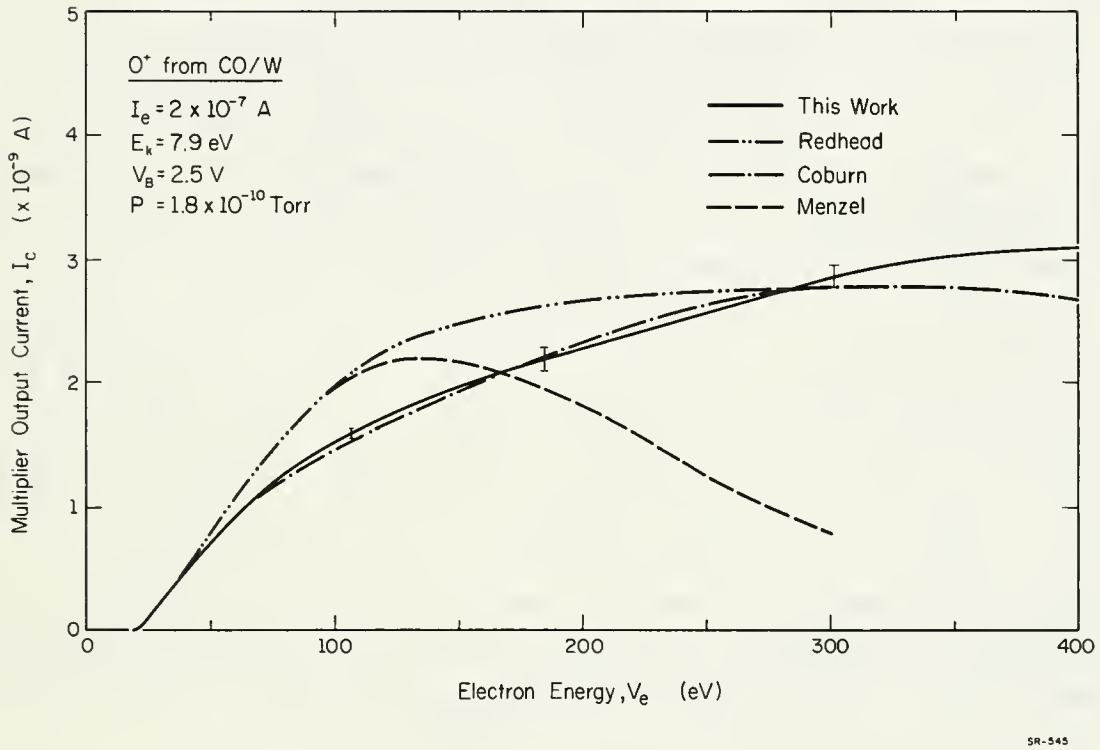


Fig. 1.10. Dependence of the O^+ ion current from CO/W on the electron energy. For comparison, data from Redhead [6], Coburn [8], and Menzel [11] are shown. The energy scale for CO/W is not corrected for the contact potential between the target and the cathode of the electron gun.

Total desorption cross sections have been measured for O_2/W, CO/W, and CO_2/W. Here we will limit our discussion only to O_2/W system. Figure 1.11 shows the variation of the total desorption cross section with electron energy. The variation of the ion desorption cross section with electron energy is also shown in the figure. The two quantities are normalized at $V_e \simeq 150$ eV. It will be noted that for $V_e \gtrsim 40$ eV, the two cross sections show a very similar dependence on electron energy, while for $V_e \lesssim 40$ eV, the curves are different, and the total desorption cross section appears to exhibit a lower threshold than that for ion desorption. It should be emphasized that the measurement of the total desorption cross section at very low electron energies is difficult and is not extremely accurate; however, the accuracy is sufficient to give strong indication of two mechanisms for the desorption of neutrals. The first mechanism appears to have a threshold at a few electron volts below the ion production threshold. Thus, this mechanism for neutral desorption is not the neutralization of ions, but probably corresponds to excitation directly to antibonding neutral states. However, since the curve for the total desorption cross section breaks at an energy close to the threshold energy for ion production and subsequently follows closely the curve for the ion desorption cross section, it would appear that the neutralization of ions is in fact also a dominant mechanism for neutral desorption.

1.1.5. Threshold Energies

The threshold energies for production of ions have been measured for O_2/W and CO/W. According to the mechanism proposed by Redhead [9], the threshold energy for the production of ions with kinetic energy, E_k, is

$$eV_i+E_d+E_k, \tag{1}$$

where V_i is the ionization potential and E_d is the activation energy for desorption of the adsorbate atom. This model assumes that both the electron incident on the system and the electron originally on the adsorbed atom are at the vacuum level immediately after the excitation process. The model proposed by Menzel and Gomer [12] assumes that one of the electrons is at the Fermi level and the other is at the vacuum level. The threshold energy given by this model is

$$eV_i+E_d+E_k-\varphi, \tag{2}$$

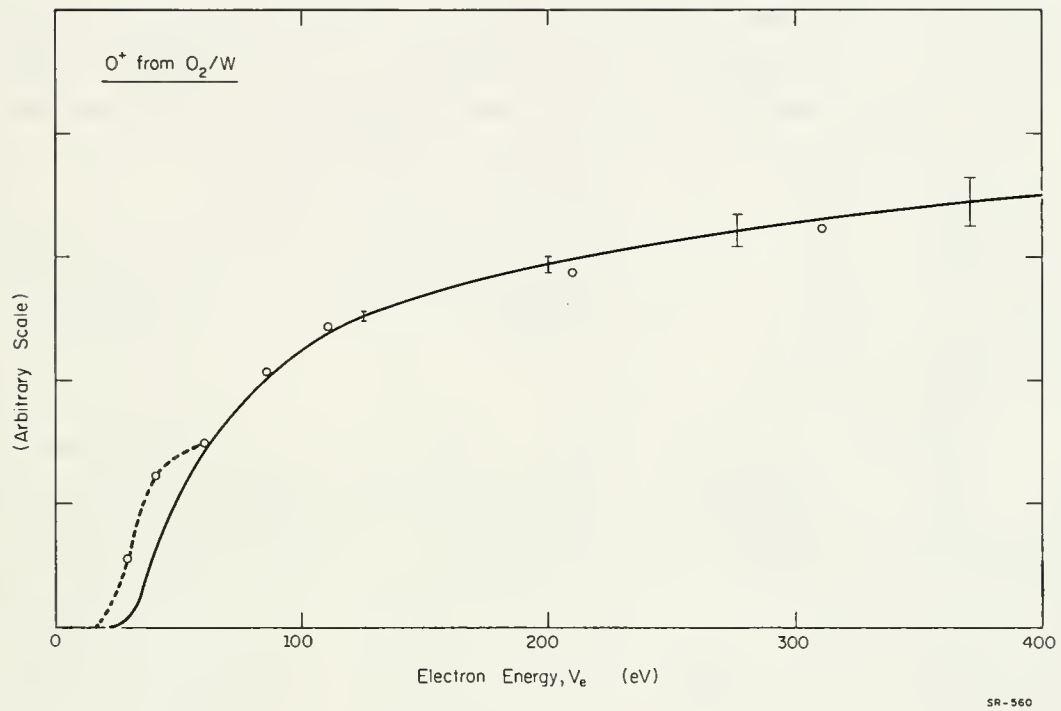


Fig. 1.11. Comparison of the variation with electron energy of the total desorption cross section (open circles) and the ion desorption cross section (solid curve) for O_2/W. The two quantities are normalized at $V_e \approx 150$ eV.

where φ is the work function of the gas covered metal. We would like to propose that the "absolute threshold" is in fact given when one assumes that both electrons go to the Fermi level, that is, that the threshold is

$$eV_i+E_d+E_k-2\varphi. \tag{3}$$

It is not necessary that this threshold be that which is measured, since there may be a very low probability for both of the electrons falling to the Fermi level. In particular, if the adsorbate is at a relatively large distance from the metal surface, the overlap integral of this case would be quite small, and thus the probability of this transition would be correspondingly small. On the other hand, if the adsorbate is quite close to the surface, we would expect that it would be more probable for both electrons to be able to go to the Fermi level. Thus, the thresholds in these two limiting cases should be close to the values given by Eqs.(1) and (3) respectively.

Figure 1.12 shows a typical threshold measurement (O^+ ions from O_2/W). Table 1.1 summarizes the threshold measurements made in the present work along with those made by other workers. The threshold obtained by other techniques have to be corrected to correlate with our results, since we measure the threshold for production of ions with a specific kinetic energy, whereas other workers typically measure the threshold for the production of ions with all energies above some low energy cutoff. It will be noted that our measurements are typically substantially lower than those previously reported. This may be due to the higher sensitivity of our apparatus. In the table, theoretical threshold energies are given for two cases: (1) the final state of the two electrons are at the vacuum level, and (2) the final states of the two electrons are at the Fermi level. The assumptions made to calculate the theoretical threshold energies are:

The work function of oxygen-covered tungsten, 5.9 eV

The work function of CO-covered tungsten, 5.3 eV

The activation energy for the desorption of an oxygen atom from oxygen-covered tungsten, 6.5 eV

The activation energy for desorption of state 1-CO from tungsten, 1.0 eV

The activation energy for the desorption of state 2-CO from tungsten, 1.1 eV

The activation energy for the desorption of a carbon atom from tungsten, 7 eV.

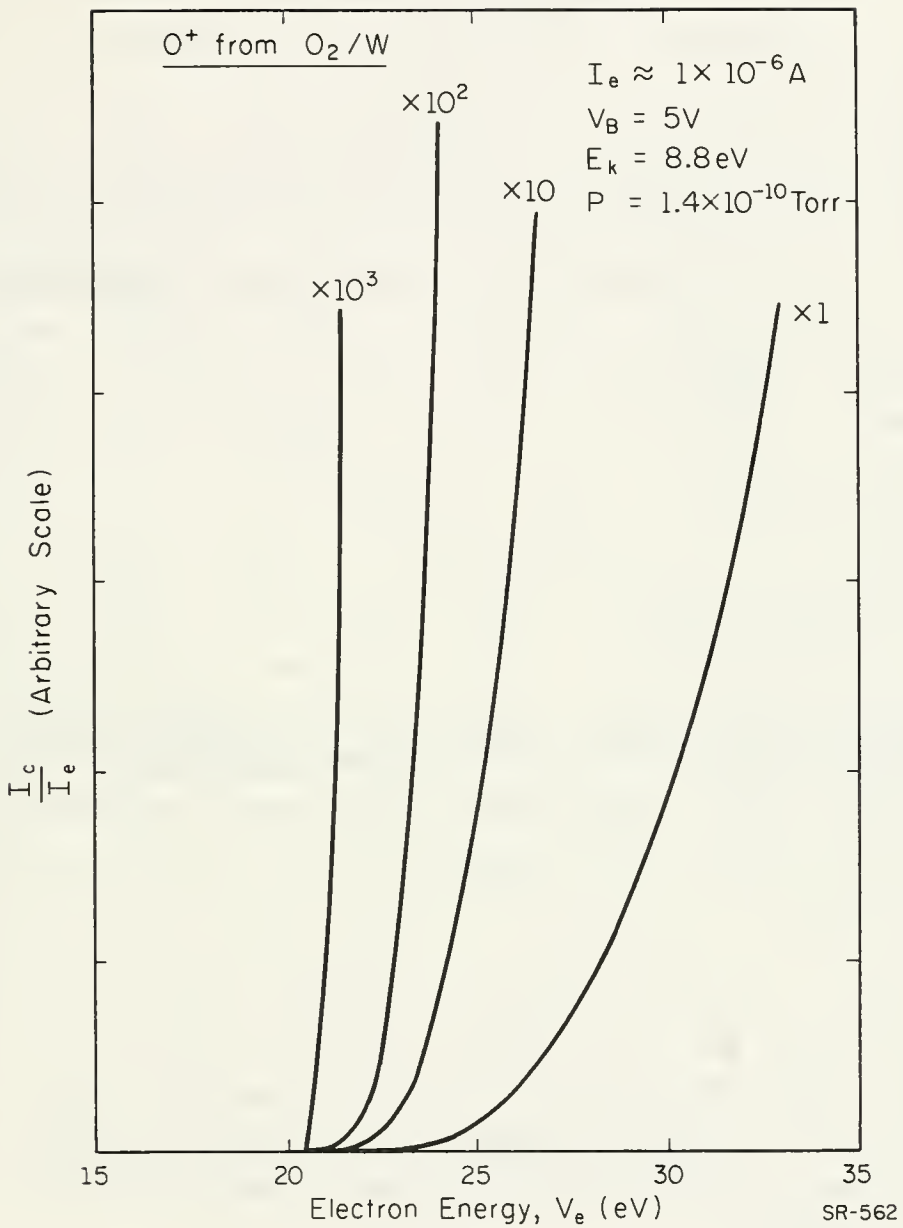


Fig. 1.12. Dependence of the O^+ ion current from O_2/W on the electron energy in the region of the threshold. The energy scale is corrected for the contact potential between the target and the cathode of the electron gun.

It is clear that in the case of O^+ ion production the measured thresholds are quite close to the values predicted when both electrons are assumed to end up at the Fermi level. In the case of CO^+ ion production, the measured threshold is quite close to the value predicted by the model in which both electrons are assumed to end up at the vacuum level. This is perhaps due to the fact that the CO molecule (in the weakly bound state which adsorbs over previously adsorbed CO) is at a substantial distance from the surface.

1.1.6. Adsorption and Desorption Measurements of CO/W System

We would like to indicate very briefly the type of result which has obtained in kinetic measurements. Figures 1.13 and 1.14 show the time dependence of the O^+ and CO^+ ion currents from CO/W during electron bombardment of the surface. The surface is initially fully covered. Figures 1.15 and 1.16 show logarithmic plots of the curves of Figs. 1.13 and 1.14. It will be noted that the curve for O^+ shows an initial increase in current with a subsequent decrease in current which is quite close to a straight line on the logarithmic plot. On the other hand, the CO^+ current can be decomposed into two straight-line portions. The length of time of the first straight-line portion of the curve corresponds approximately to the time in which the O^+ current is passing through a maximum. These results are interpreted to indicate that there are two states of adsorption active. The first state has a relatively large cross section for the production of CO^+ and an even larger cross section for electron conversion to the second state. This second state has a relatively large cross section for O^+ ion production and a much smaller (1/40) cross section for CO^+ ion production. That there is one state which produces both types of ions is indicated by the fact that the time constant (and thus the total desorption cross section) for the O^+ curve is the same as that for the second portion of the CO^+ curve.

Interesting conversion effects have also been observed in other systems, especially with H_2O and CO_2.

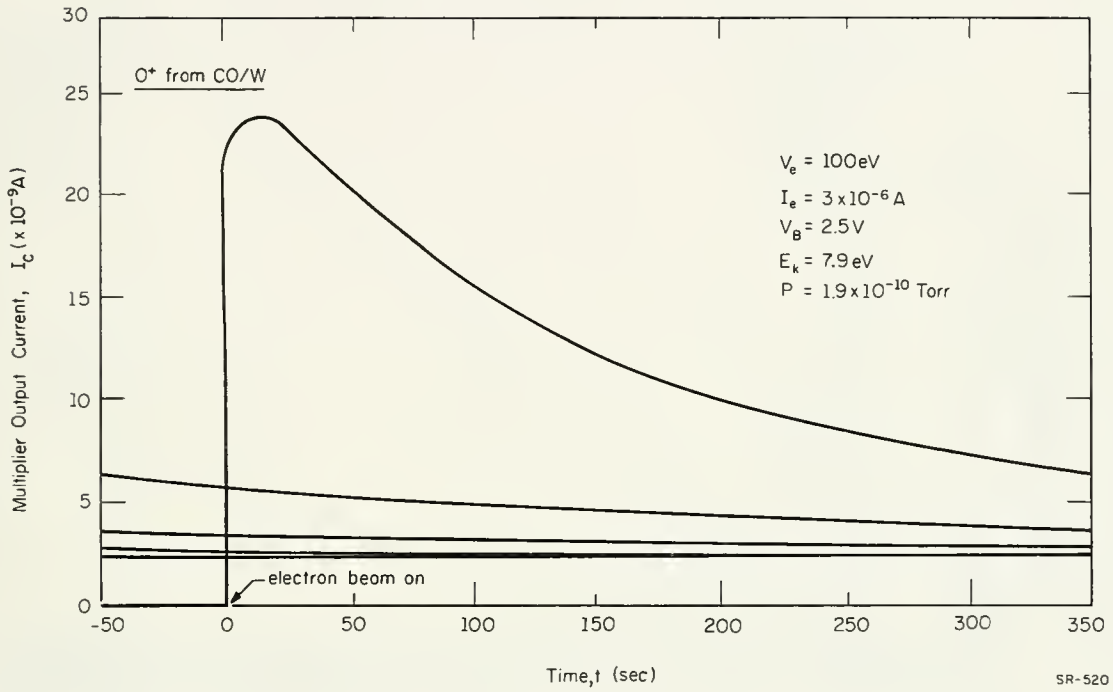


Fig. 1.13. Change of the O^+ ion current during electron bombardment of tungsten initially fully covered with CO.

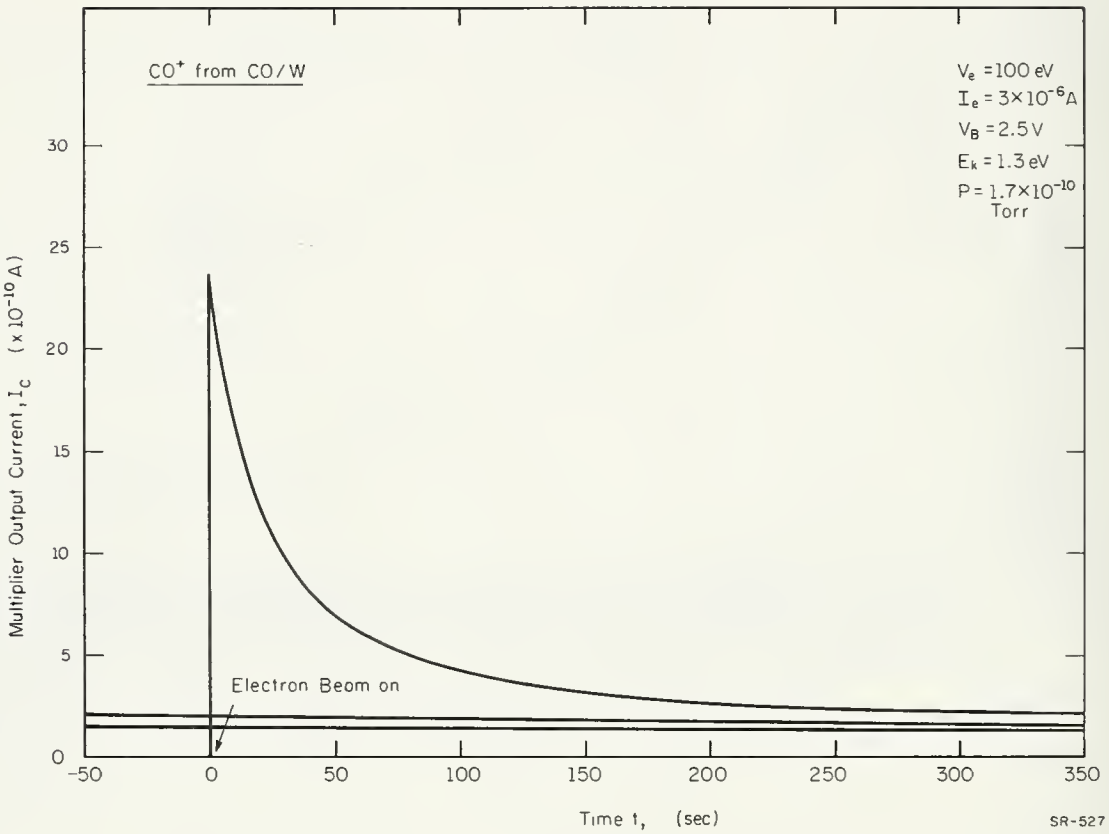


Fig. 1.14. Change of the CO^+ ion current during the electron bombardment of tungsten initially fully covered with CO.

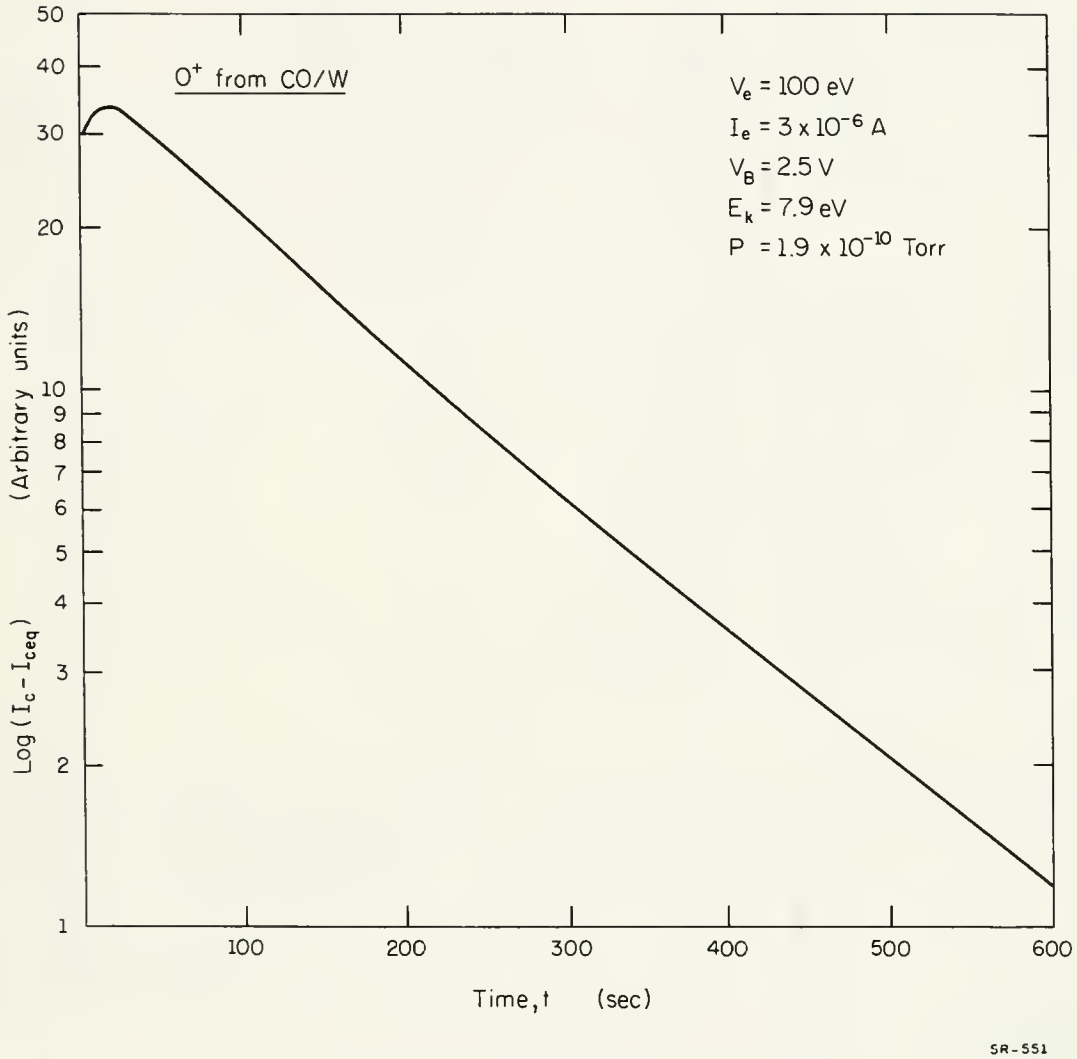


Fig. 1.15. Semilogarithmic plot of the O^+ ion current as CO desorbs from the tungsten surface by electron bombardment. The data plotted are those from Fig. 1.13.

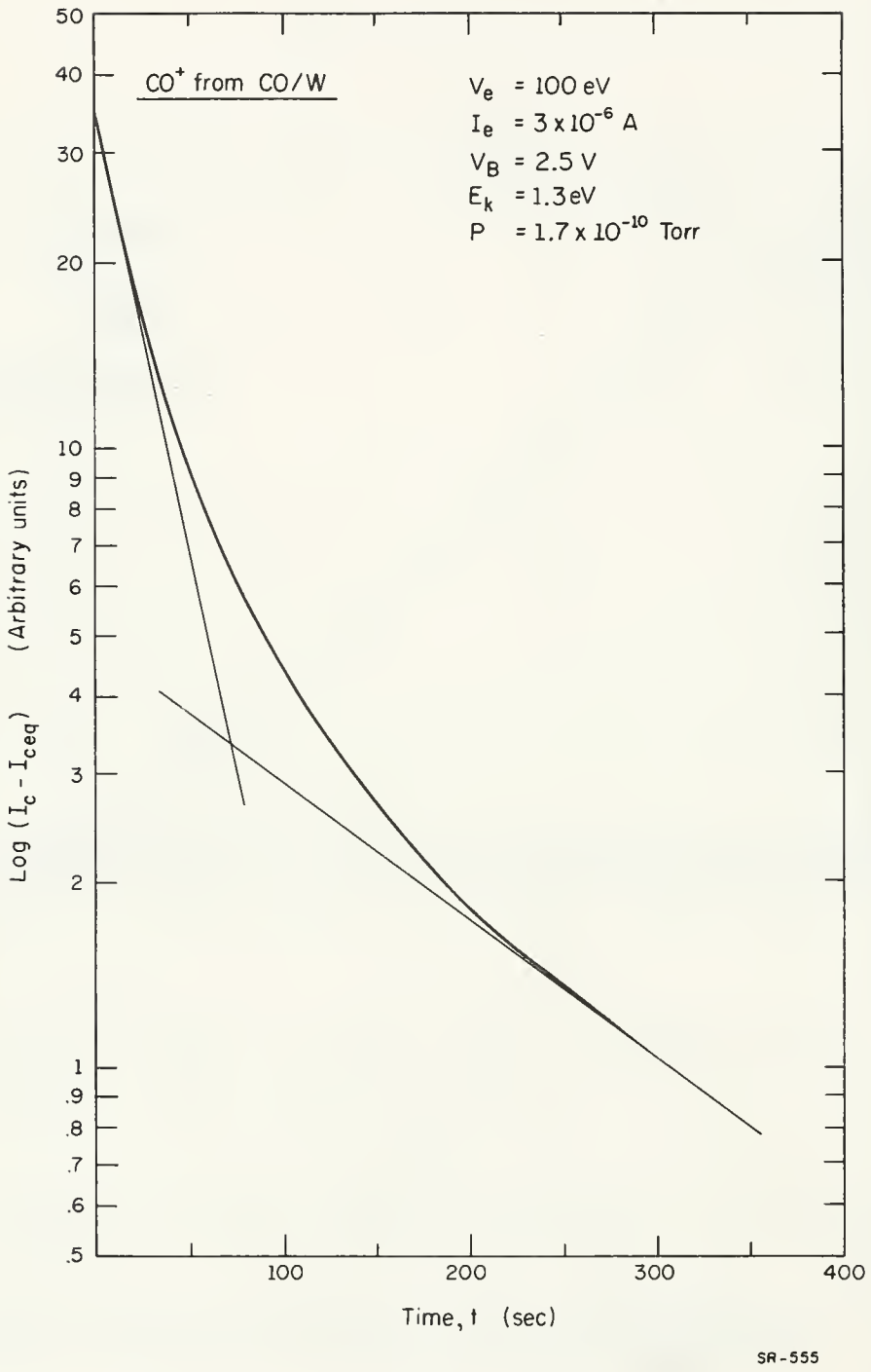


Fig. 1.16. Semilogarithmic plot of the CO^+ ion current as CO desorbs from the tungsten by electron bombardment. The data plotted are those from Fig. 1.14.

Table 1.1. Comparison of the Threshold Energy of Ion Desorption Obtained by Various Workers.

System	Ion	Worker	"Min. Threshold Energy"(eV)*	Norm. Results (eV)	Theoretical Threshold Energy (eV),** Final States of the Two Electrons at Vacuum Level	Fermi Level
O_2/W	O^+	Yates et al.[13]	21.8(E_k = ~5)	~25.6(E_k=8.8)	28.9	17.1
		Madey, Yates [7]	19.3(E_k = ~5)	~23.1(E_k=8.8)	(E_k=8.8eV)	(E_k=8.8)
		This Work	20.5(E_k=8.8)†	20.5(E_k=8.8)		
CO/W	O^+	Redhead [6]	20.5(E_k = ~4)	~24.4(E_k=7.9)	26.7	16.1
		Yates et al.[13]	17.3(E_k = ~4)	~21.2(E_k=7.9)	(E_k=7.9)	(E_k=7.9)
		Coburn [8]	20.9(E_k=4.7)†	24.1(E_k=7.9)		
		Menzel [11]	20.0(E_k = ~4)	~23.9(E_k=7.9)		
		This Work	18.5(E_k=7.9)†	18.5(E_k=7.9)		
CO/W	CO^+	Redhead [6]	15.1(E_k = ~0)	~16.4(E_k=1.3)	16.4	5.8
		Yates et al.[13]	17.3(E_k = ~0)	~18.6(E_k=1.3)	(E_k=1.3)	(E_i=1.3)
		Menzel [11]	14.6(E_k = ~0)	~15.9(E_k=1.3)		
		This Work	15.5(E_k=1.3)†	15.5(E_k=1.3)		

*The "minimum threshold energy" is the threshold for the production of ions with kinetic energy equal to the low energy cutoff of the ion energy distribution. The kinetic energy of the corresponding ions is an estimate based on the energy distributions measured in this work.

†The threshold energy for these measurements is for ions of the kinetic energy specified.

**Assumptions used in this calculation are given in the text.

F. Propst
C. Foster
M. Nishijima

1.2. Energy-Angular Distributions of Secondary Electrons‡

During this reporting period, a considerable amount of data has been taken, several important apparatus modifications have been made, and a better understanding of some of the experimental difficulties has been developed.

‡Supported principally by the Joint Services Electronics Program (U.S. Army, U.S. Navy, U.S. Air Force) under contract DAAB-07-67-C-0199 and in part by the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

These developments will be discussed briefly here, and in more detail in an R report now being prepared.

The operation of the titanium sublimation pump reported previously [14] indicated that the background pressure is made up of a large fraction of a nongetterable gas which is almost certainly mercury vapor. Partial pressure measurements reported earlier [14] indicate that the pressure rise of about 30 to 40 percent as the getter becomes saturated is due to hydrogen diffusing from the stainless-steel components and chamber walls.

It was found that the 25-μm diameter molybdenum wires that formed the analyzer grids [15] had become "dead soft" during the brazing process. It was decided to attempt to construct the grids from tungsten wire. From 12.5μm tungsten wire, much more rugged grids were easily made. The smaller wire allowed a smaller grid wire spacing without decreasing the transparency (90%), and thus better analyzer performance.

Considerable effort was expended in the study of alignment problems. Deflection units were added to the electron gun. With these, it was found that at least some of the anomolous intensity variation [15] in the diffraction patterns was due to misalignment.

The construction of each component was carefully checked for misalignment and, where necessary, adjustments were made. Fixtures for aligning the components relative to one another were constructed. The improved alignment, along with improved target preparation, proved to be the most important system modifications.

Figure 1.17 shows an energy distribution of secondary electrons for a primary energy of 114 eV. The positions of the peaks are in reasonable agreement with those reported by other investigators [16], including the 5 eV loss peak previously detected on the high resolution spectrometer [17] in this laboratory. The 5 eV loss peak is not nearly so predominant at other angles and primary energies as it is in the energy distribution shown in Fig. 1.17. In many instances, in fact, its presence is only detectable with the prior knowledge of its existence and its gas coverage properties. The most predominant peaks are usually the ones located at about 10 eV and 24 eV in Fig. 1.17.

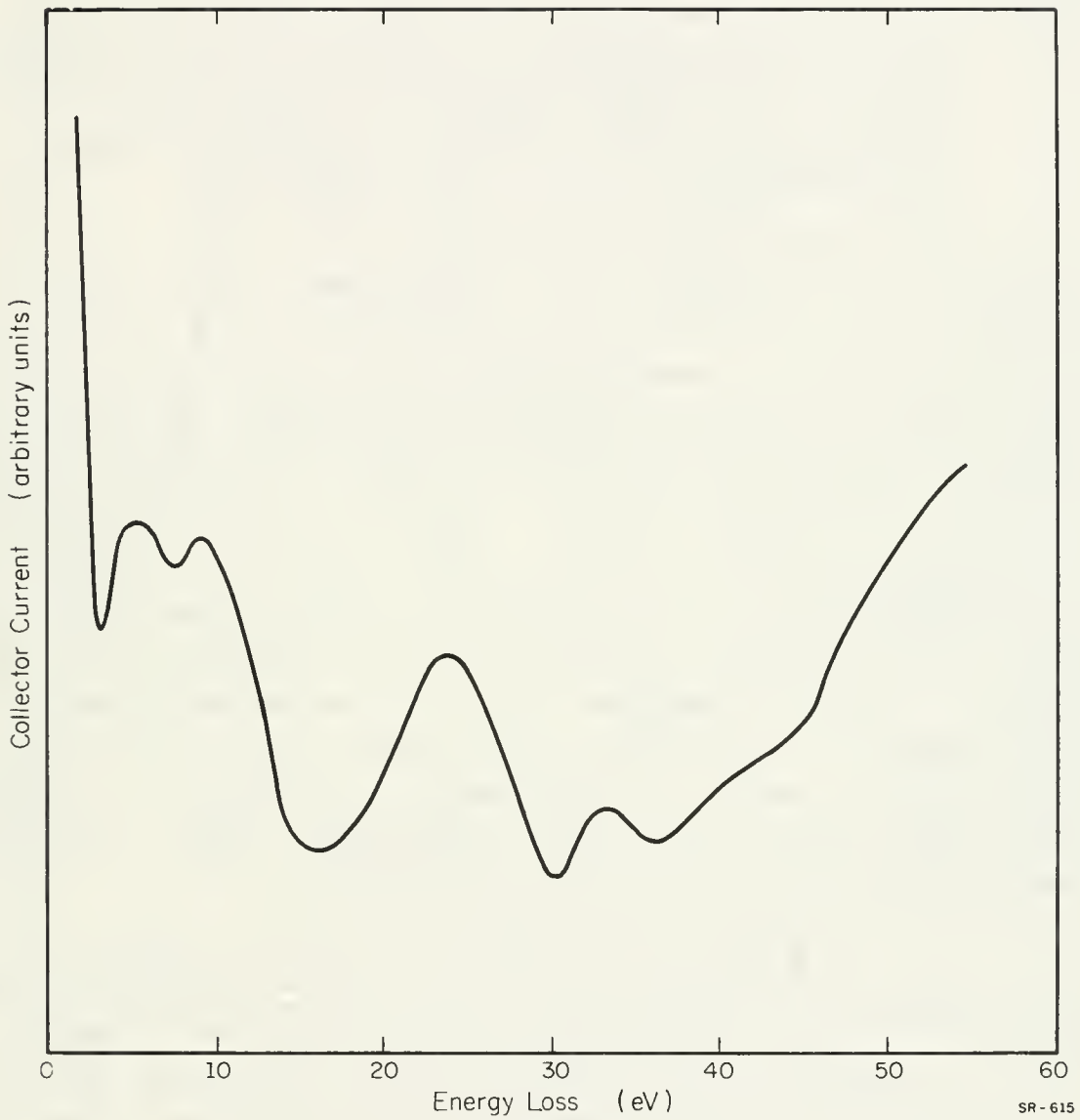


Fig. 1.17. Energy distribution of electrons scattered from clean 100 tungsten resulting from the bombardment of 89 eV electrons.

One of the most striking features of the energy distributions has been the apparent shifts in the energy positions of the two predominant characteristic loss peaks. This effect is illustrated in Fig. 1.18. Similar shifts have been observed with the high-resolution spectrometer.

The angular distributions have produced three important results

structure in the angular distribution of inelastically scattered electrons which correlate with diffraction peaks,

shifts of these peaks from the diffraction angles,

splitting of the inelastic peaks.

These effects are illustrated by the curves in Figs. 1.19, 1.20 and 1.21. The angular distribution for elastic reflection shown in Fig. 1.19 exhibits only one peak, whereas the inelastic distributions show two strong peaks. Because of misalignment of the system components, the target must be rotated a few degrees in order for the other elastic peak to be detected. With this proviso, the correlations of both the 10 eV and 24 eV loss peaks with the diffraction peaks is evident. Whether the splitting in the 24 eV loss distribution is a real physical effect or due to systematic experimental error is a question of some importance, and has been under intensive consideration. Since such splittings occur at diffraction angles, and the elastic component of the beam entering the detector is about 100 times the detected inelastic current, any ac disturbance (at the modulation frequency) of the elastic component will contribute an erroneous signal. We have some evidence that this type of error does, in fact, exist.

Figure 1.20 shows the angular distribution of another 10 eV loss peak. This curve also illustrates splitting. However, the minimum between the two peaks for this curve does not correspond to the diffraction angle of the elastic beam but to the diffraction angle for an incident beam with an energy equal to the secondary energy of the 10 eV loss peak. Therefore, the splitting of the inelastically scattered electrons cannot be due to an extraneous ac signal on the elastic component of the detected beam.

At this time, however, it is not possible to say with any real certainty to what extent splitting in many curves, such as the 24 eV loss distribution in Fig. 1.19, is due to experimental effects and to what extent it represents a real physical phenomenon.

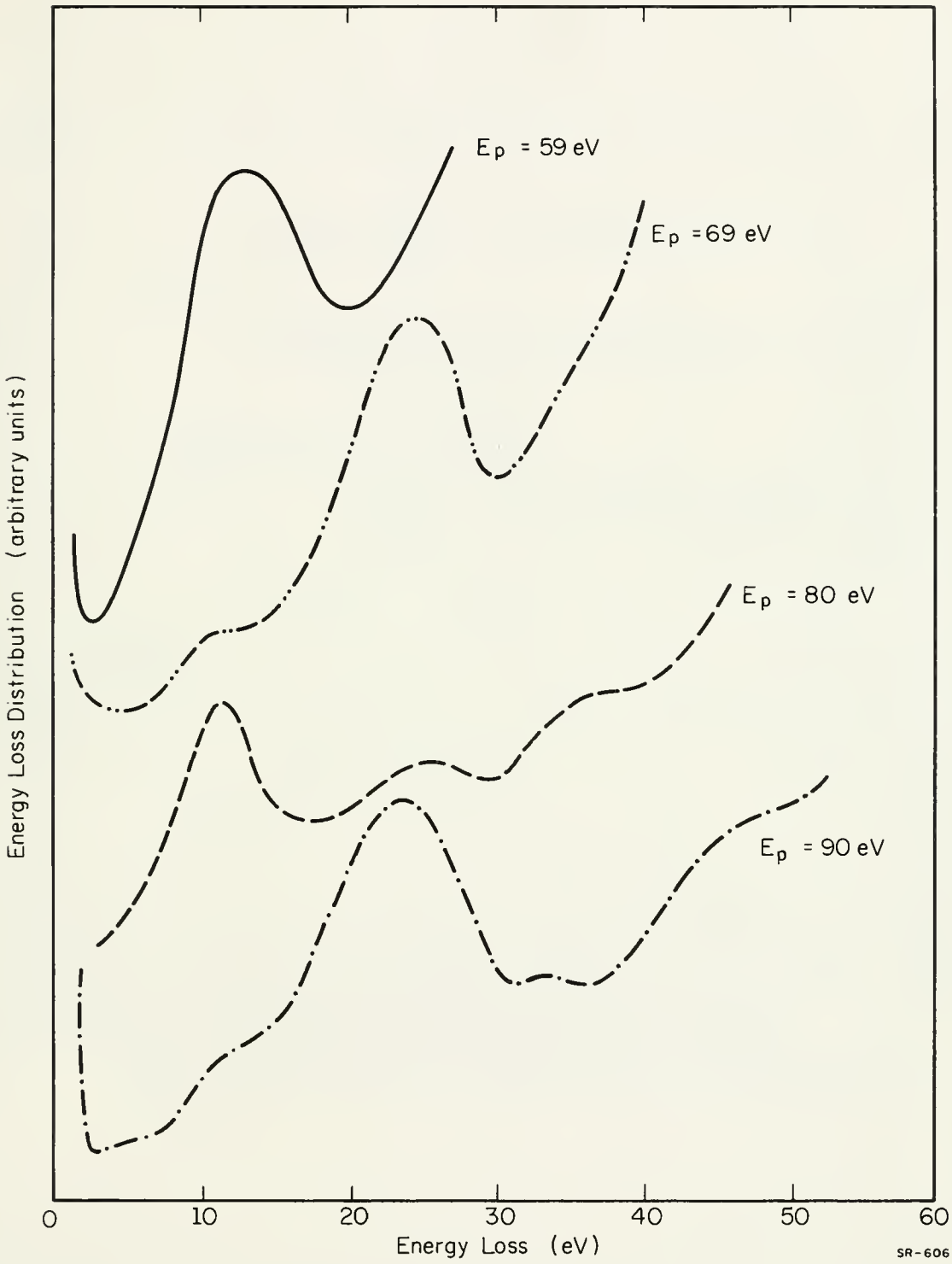


Fig. 1.18. Energy distributions of electrons scattered from clean 100 tungsten resulting from the bombardment of 59, 69, 80 and 90 eV electrons.

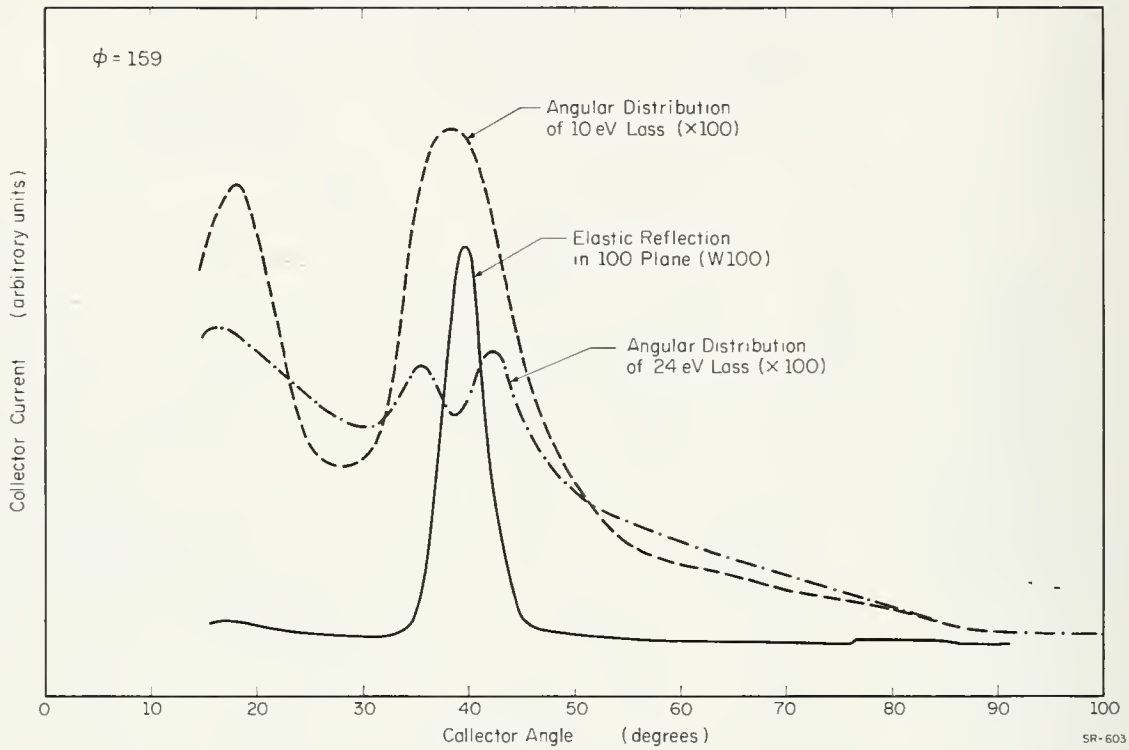


Fig. 1.19. Angular distributions of electrons scattered from clean 100 tungsten resulting from the bombardment of 114 eV electrons.

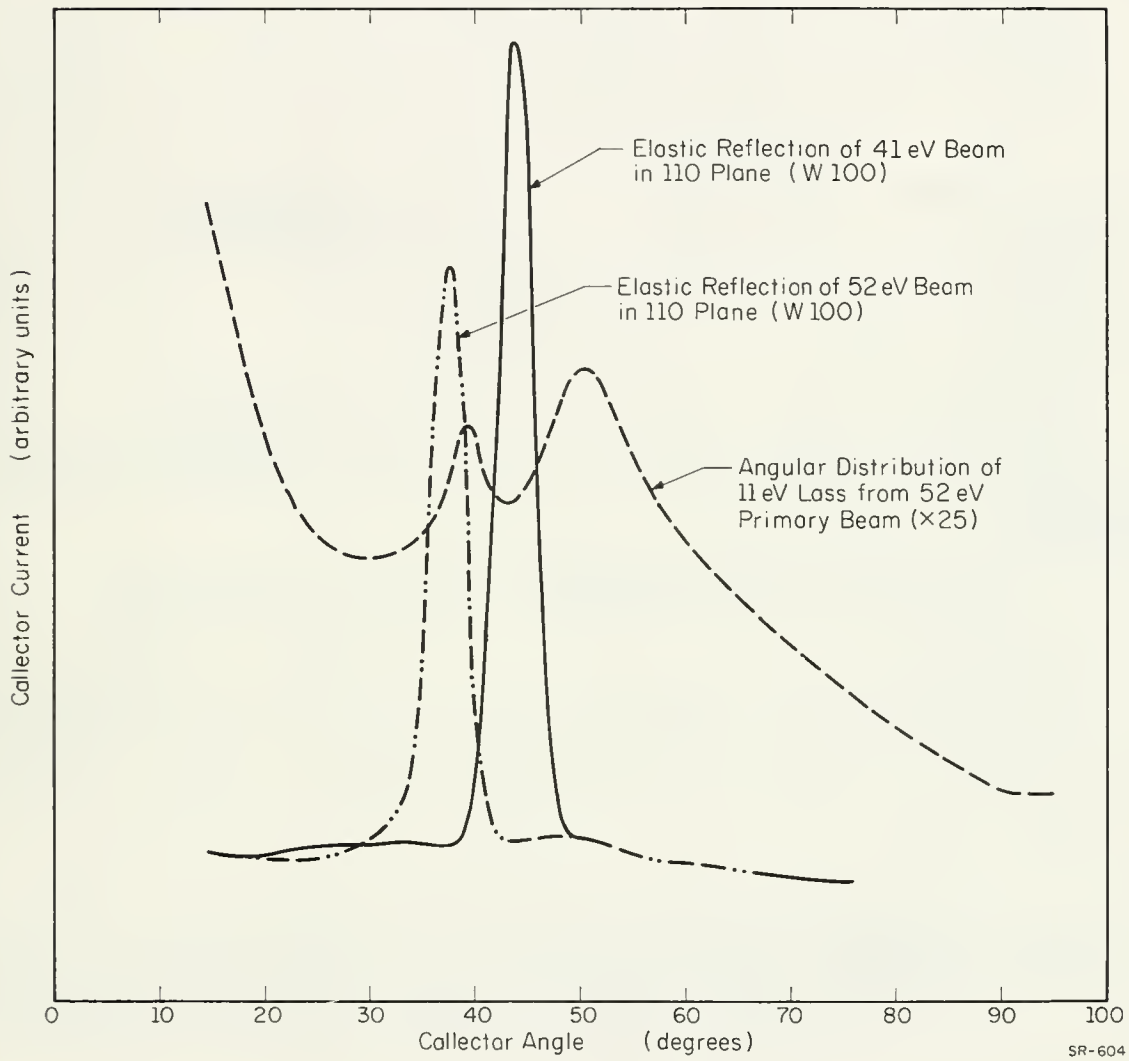


Fig. 1.20. Angular distributions of electrons scattered from clean 100 tungsten resulting from the bombardment of 52 and 41 eV electrons.

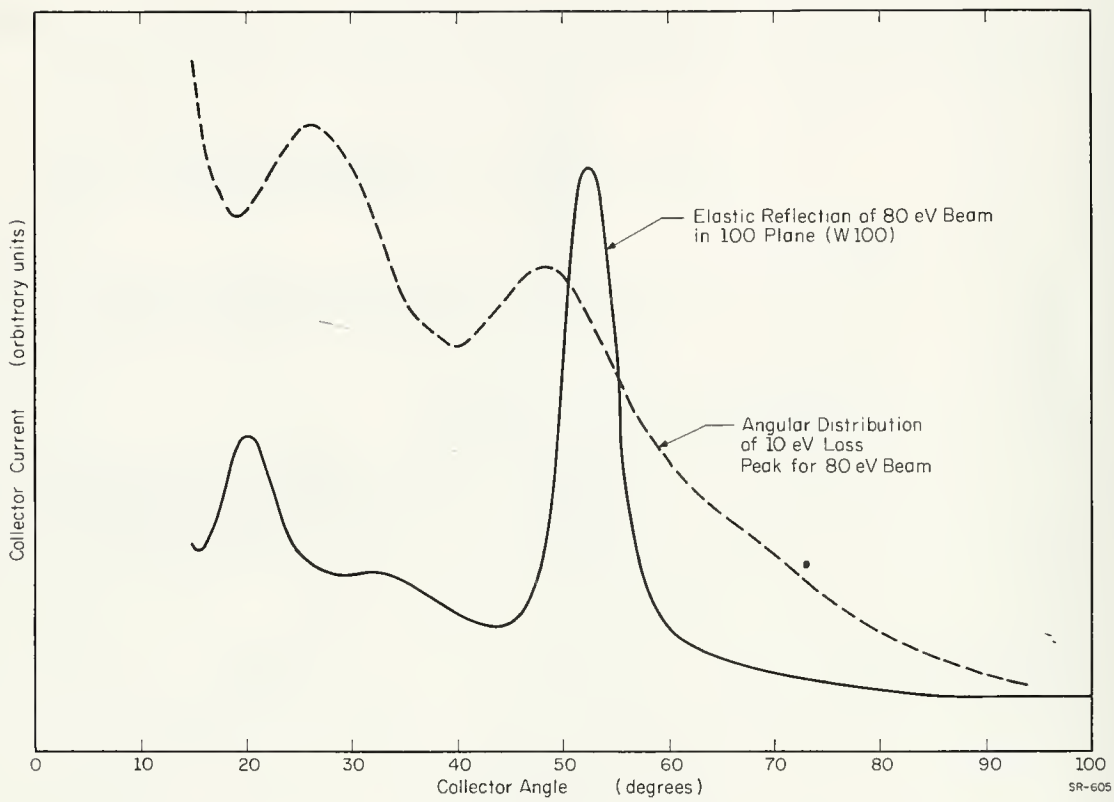


Fig. 1.21. Angular distributions of electrons scattered from clean 100 tungsten resulting from the bombardment of 80 eV electrons.

Figure 1.21 illustrates shifts of the inelastic angular peaks away from the diffraction angles. Shifts such as these contain information concerning the momentum transfer during the inelastic collision.

One other area which has been examined briefly is the area of gas adsorption. Figure 1.22 shows four energy distributions taken after the target was allowed to adsorb background gas for several hours. The corresponding clean target curves are those in Fig. 1.18.

The apparatus, because of its construction, had remained in the Engineering Research Laboratory after the completion of the New Coordinated Science Laboratory. Near the end of this report period, a failure in the apparatus occurred. It was decided at this point to dismantle the apparatus and move it to the new laboratory. It is now in the process of being reassembled. When the equipment is once again operational, it is planned to undertake a detailed study of some of the effects reported here.

F. Propst
J. Burkstrand
T. Cooper*

1.3. High-Resolution Electron Spectroscopy of Solid Surfaces*

Using the high-resolution electron spectrometer described in previous reports, we have measured the elastic-reflection coefficient and secondary-energy distribution for electrons scattered from the 100 surface of a tungsten crystal. A summary of the experimental findings will be given here; the entire experiment will be reported in an R report.

In the experiment the beam-crystal orientation angle φ (see Fig. 1.23) was set to 0°, 22.5°, 45° and in each configuration the elastic-reflection coefficient and secondary electron distribution were measured. Figures 1.24, 1.25 and 1.26 show the elastic-reflection coefficient which we measured at the three angular positions.

In view of the recent theoretical calculations of Duke, Tucker and Duke, Anderson, Tucker (DAT), it is possible to tentatively identify the structural forms seen in the elastic reflection coefficient at $\varphi = 0^{\circ}$.

*Supported principally by the Petroleum Research Fund of the American Chemical Society, PRF3668-3,5 and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199; and in part by the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

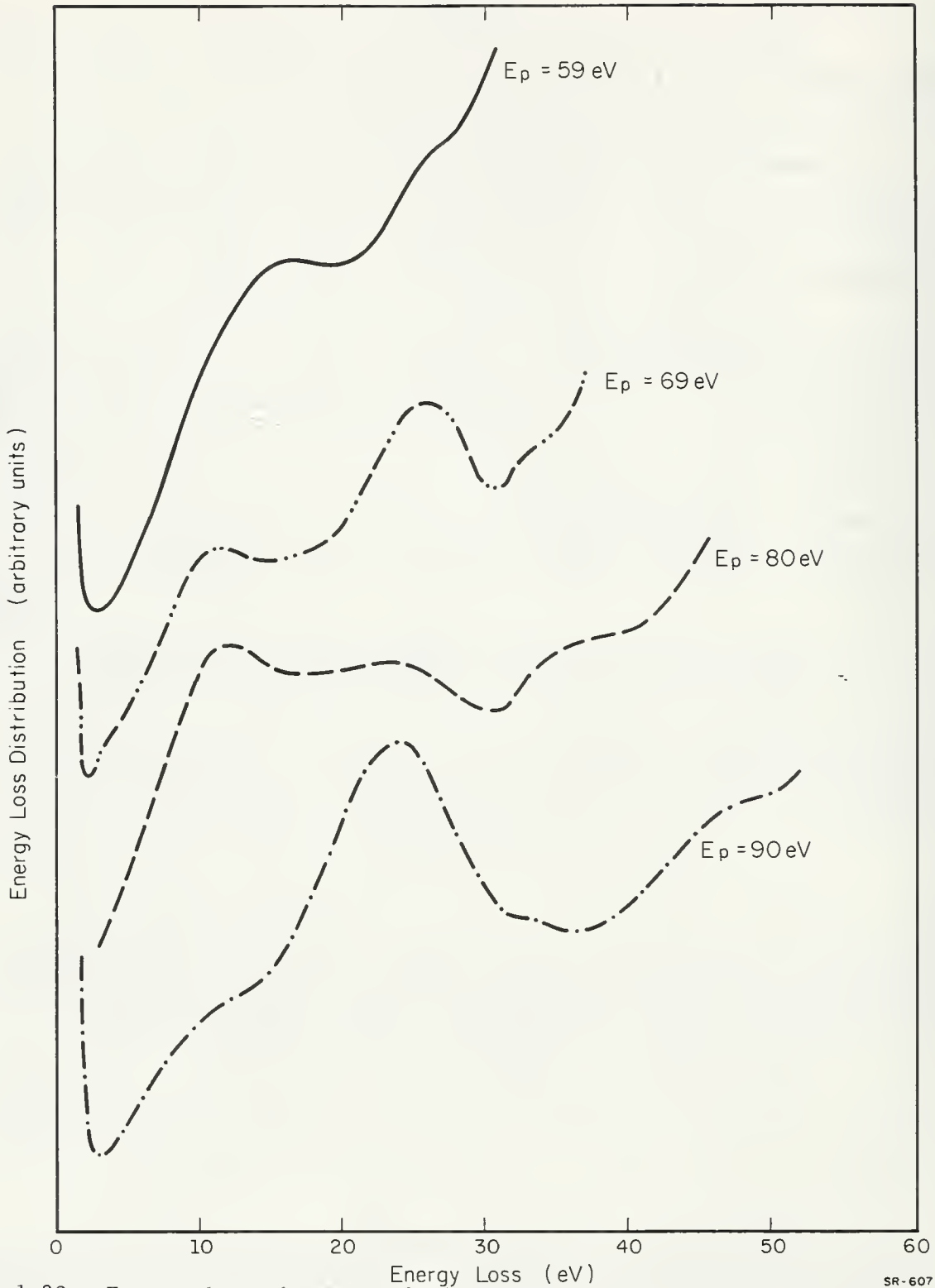


Fig. 1.22. Energy distributions of electrons scattered from gas covered 100 tungsten due to the bombardment of 59, 69, 80, and 90 eV electrons.

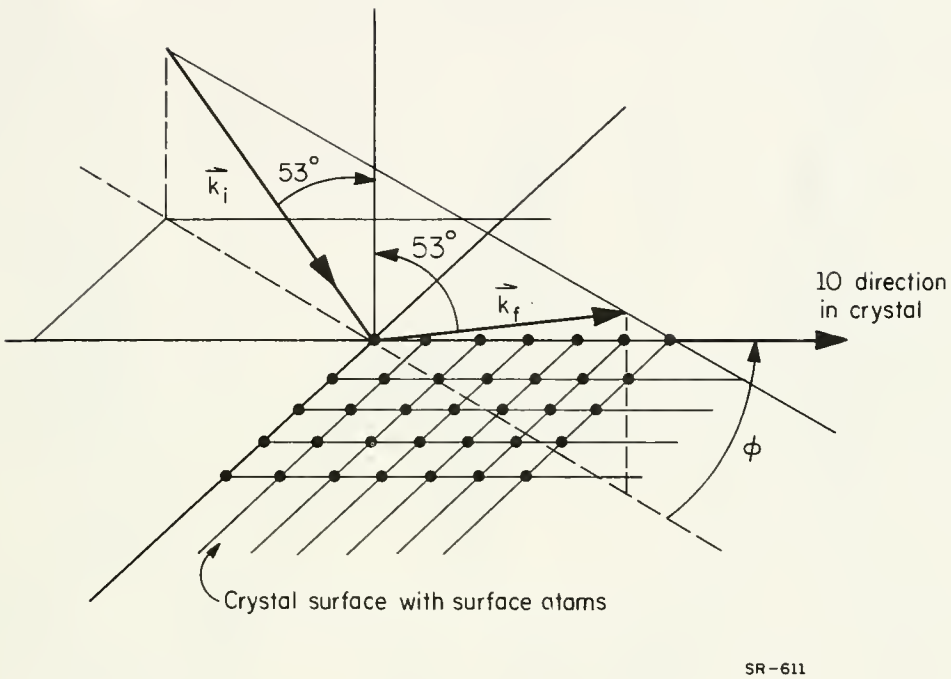


Fig. 1.23. Diagramatic illustration of the incident and final electron beam directions with respect to the crystal-surface orientation.

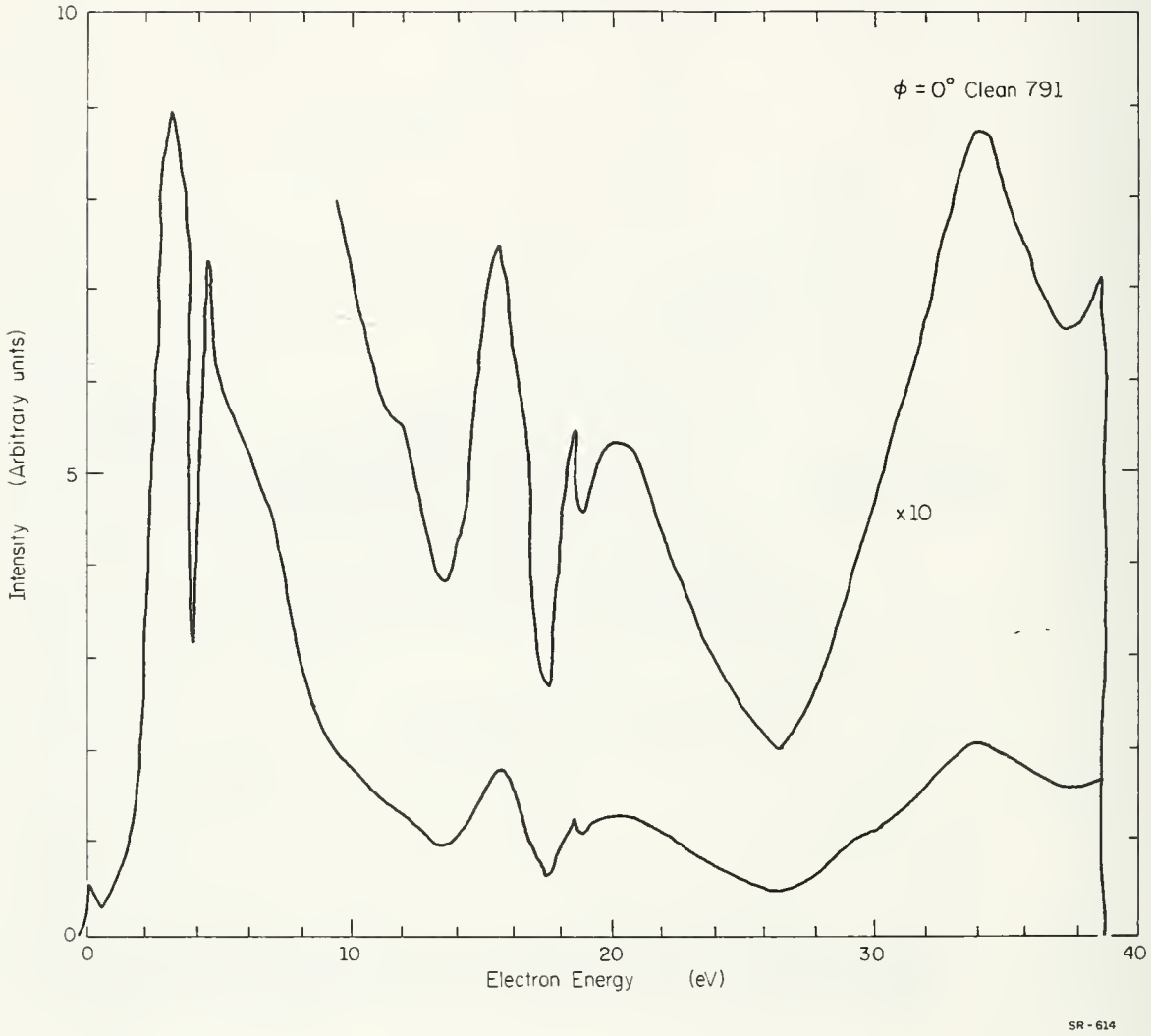


Fig. 1.24. Elastic reflection coefficient as a function of electron energy for electrons scattered from a clean W(100) surface with $\varphi = 0^{o}$.

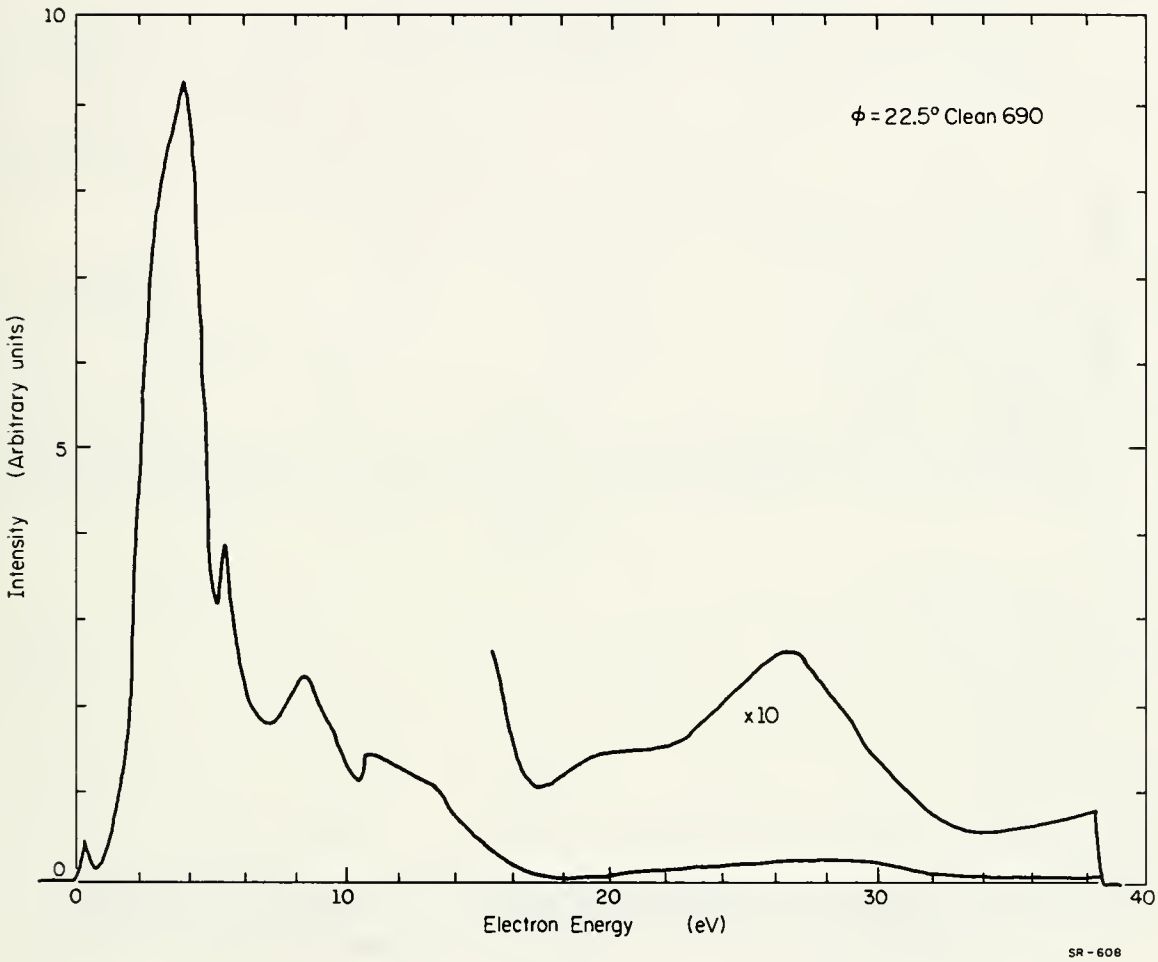


Fig. 1.25. Elastic reflection coefficient as a function of electron energy for electrons scattered from a clean W(100) surface with $\varphi = 22.5^{\circ}$.

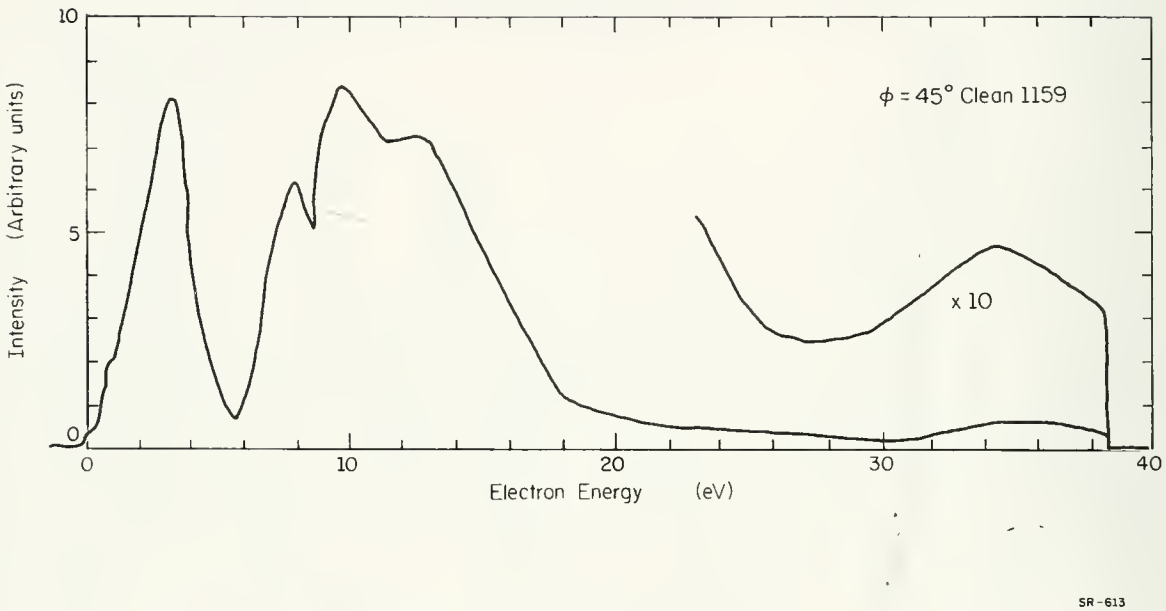


Fig. 1.26. Elastic reflection coefficient as a function of electron energy for electrons scattered from a clean W(100) surface with $\varphi = 45^{\circ}$.

On comparing our data with a calculation made by Duke, Anderson, Tucker it appears that the 3-4 eV structure is a surface resonance, the 15-22 eV structure secondary Bragg diffraction and the 30 eV structure primary Bragg diffraction. Of extreme interest and potential excitement is the low value of inner potential (4 eV) used by DAT to match our data as opposed to band theory predictions of 11 eV inner potential. It should be mentioned that a completely unambiguous interpretation of the respective structural phenomena in the elastic reflection coefficient will be possible only after a high-resolution angular experiment is performed.

The sensitivity of the interpreted surface resonance may be seen by noticing the change in the structure in the elastic reflection coefficient under brief oxygen exposures. Notice in Fig. 1.27 the rapid decrease of the 3.5-4.5 eV structure during O_2 adsorption while the remaining structure remains relatively constant and unaffected. Thus, we see in a very clear way that the 3-4.5 eV structure depends much more critically on surface order than does the remaining structure and complements the interpretation of the phenomena on the basis of DAT.

Apart from the elastic reflection coefficient, we have measured the secondary electron distribution of electrons scattered from our crystal. Two types of phenomena appear in this measurement. The first are peaks which are located at a constant energy from the elastic peak. The second are peaks which are located at a constant energy from the zero of secondary energy. The peaks located a constant energy from the elastic peak are due to discrete loss mechanisms of the electron-metal system. These data are shown in Fig. 1.28. The 10 eV and 23.5 eV losses, seen also by Scheibner and Tharp on W(110) are interpreted as being, respectively, surface and bulk plasmon excitations; the bulk excitation energy being 22.5 eV and the surface excitation energy being 16.5 eV.

Also seen in our data are loss peaks at 1.5, 4.2, 33, 43, 53 and possibly one at 16 eV. These, as far as we know, are due to separate interband transitions, but do not seem to correspond to any predicted interband transitions. More study is obviously needed in this area to be able to extract more information from experimental measurements.

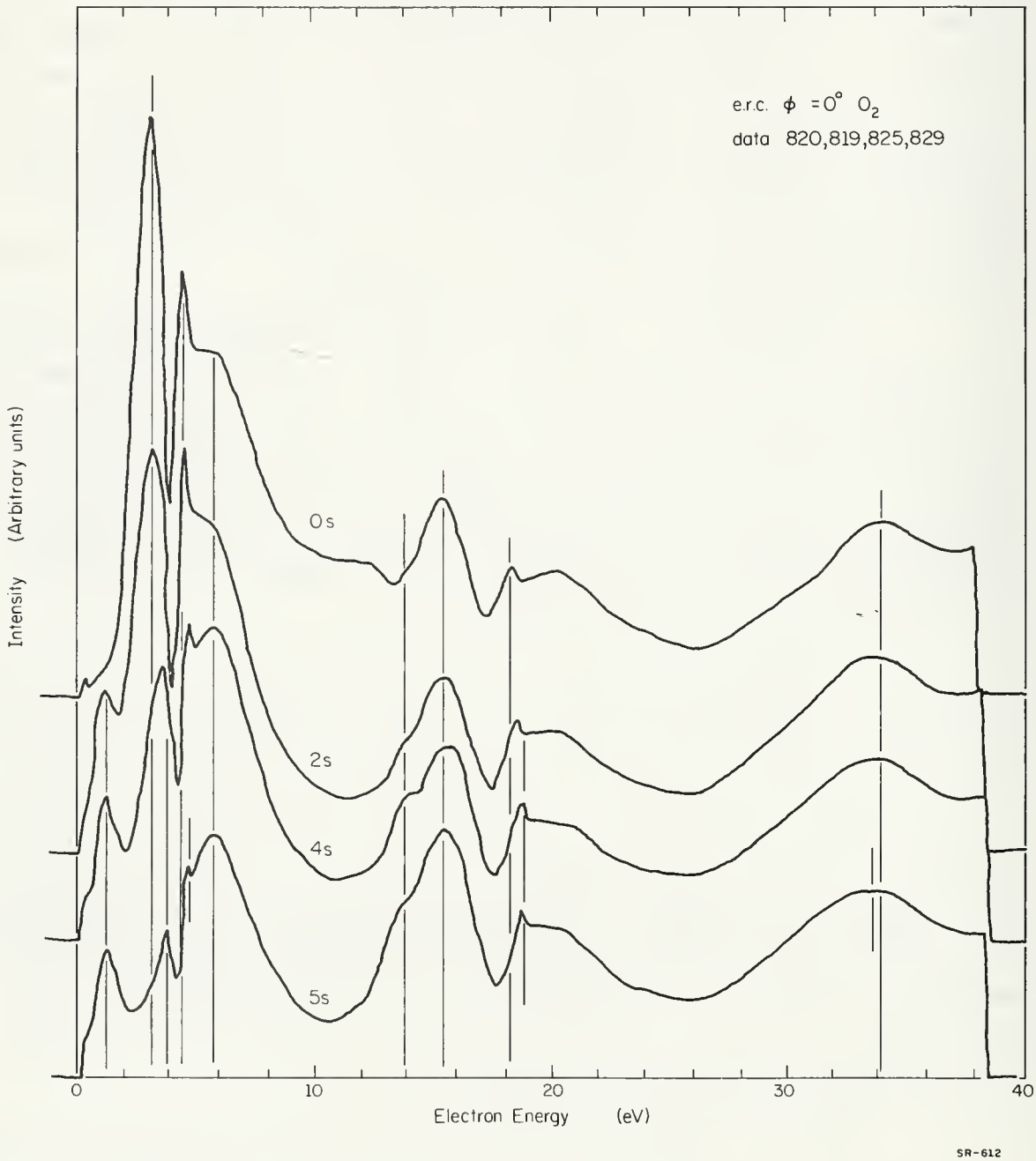


Fig. 1.27. Elastic reflection coefficient as a function of electron energy for various exposures to oxygen.

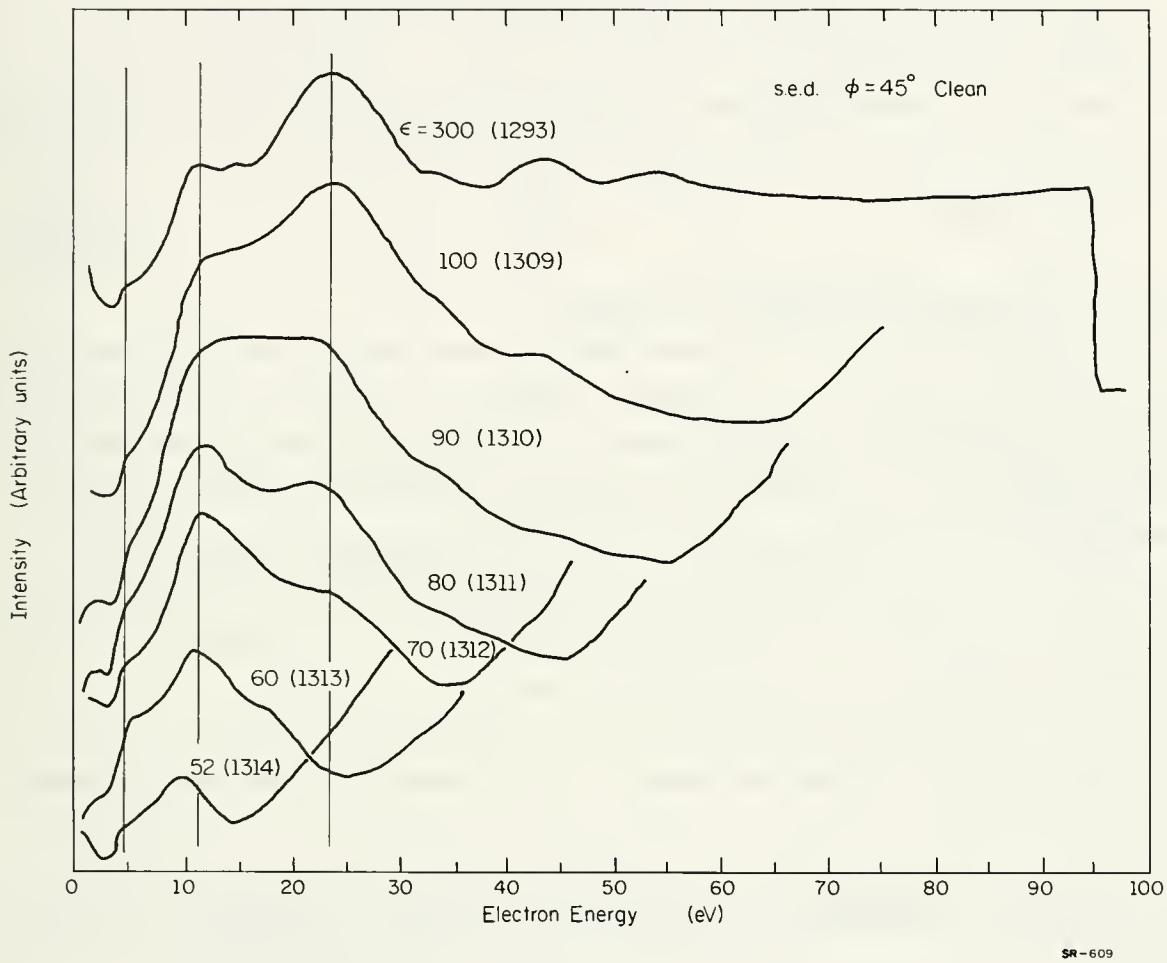


Fig. 1.28. Secondary electron distribution as a function of energy loss for various primary energies.

The peaks located at a constant energy from the zero of secondary energy, independent of the incident electron energy are called secondary energy peaks and are shown in Fig. 1.29. These peaks have been interpreted as being due to electrons excited in Auger transitions. We believe that, in addition to this possibility, the interpretations based on a model initially proposed by Davison and Germer may be possible. This model predicts a correlation between the peak positions in the elastic reflection coefficient and secondary electron distribution. This will be discussed in the subsequent R report.

F. Propst
D. Edwards

1.4. Chemical Dynamics on Solids‡

During the last decade, modern experimental techniques and equipment have progressed to the stage where the preparation and maintenance of crystal surfaces in a well-defined state is now a matter of routine. This has for the first time made it possible to explore the chemistry of the gas-solid interface on a level approaching the atomic. As a consequence, there is hope that the events involved in a variety of technologically important surface processes, such as oxidation or heterogeneous catalysis, can be established on a sound foundation.

This program of studies was initiated in the fall of 1969 to take advantage of these developments and is aimed at achieving an understanding of the atomistics of gas-solid interactions. The philosophy is to subject surface problems to coordinated examination by a variety of modern physical techniques capable of visualizing the behavior of gases on crystals. The work is still in its early stages and is now largely involved with building up the necessary experimental facilities. The areas to be covered, as well as the present state of development of the projects, are briefly sketched below.

1.4.1. Surface Specificity in Adsorption

Chemical reactions occurring at surfaces must in some way involve

‡Supported by the U.S. Air Force Office of Scientific Research of the Office of Aerospace Research under Contract Number AFOSR 69-1671 and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199.

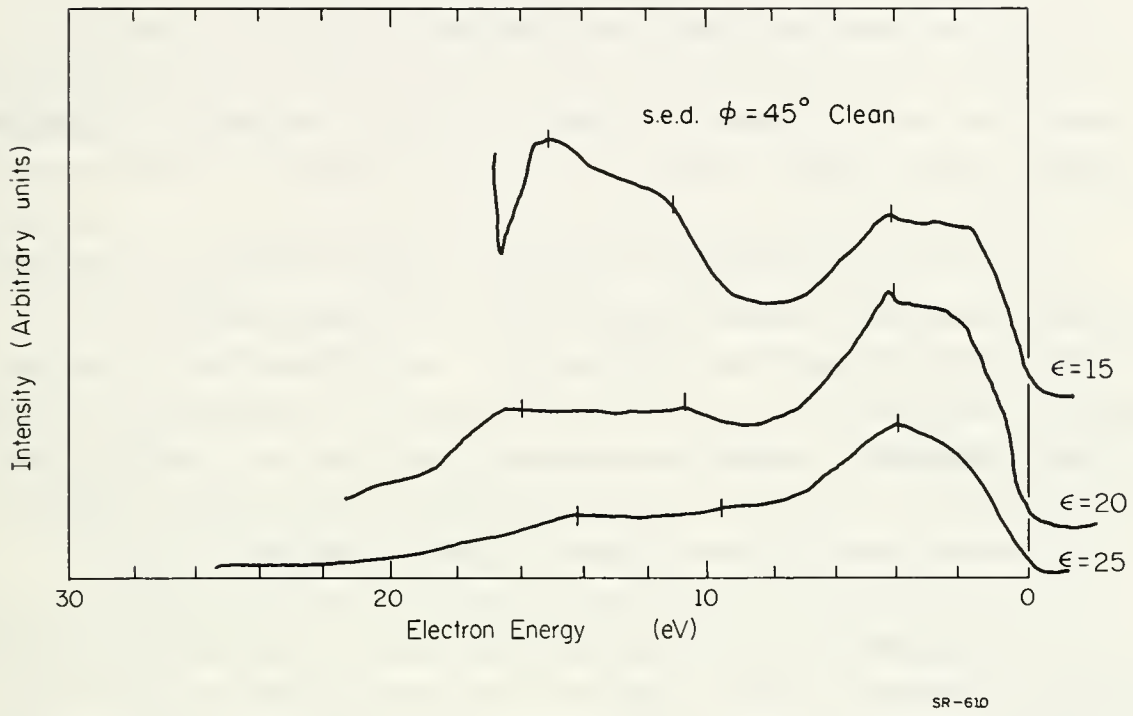


Fig. 1.29. Secondary electron distribution as a function of electron energy for $\varphi = 45^{\circ}$.

adsorbed entities. An understanding of such reactions therefore hinges upon our ability to specify the nature and properties of this adsorbed layer. In the past, it has been demonstrated that the atomic arrangement of the surface, that is, the geometry of the crystal planes exposed, is vital in affecting the physical and chemical properties of gases held at the surface. As one part of the program, we therefore plan to examine the behavior of simple adsorbed gases as a function of surface crystallography. To this end adsorption studies using the field emission microscope have been initiated [19].

In this microscope, an image of a small hemispherical emitter is projected onto a fluorescent screen by electrons tunneling out of the emitter under the influence of a high electric field (3GV/m). The image maps the electron emission from the surface with a resolution on the order of 2nm. Inasmuch as adsorption changes the electron work function, the field-emission microscope allows us to follow adsorption as it occurs on the different crystal planes forming the emitted surface.

During this reporting period, an ultrahigh-vacuum system expressly designed for field-emission microscopy has been installed. High-voltage equipment, as well as precision current and voltage detectors, have been acquired. Work on the determination of the work function of individual planes will commence shortly. Initially, this will be concentrated on studies of simple inert gases such as xenon adsorbed on tungsten, in order to provide a background for later, more chemically-oriented, studies.

1.4.2. Molecular Dissociation on Crystals

Any heterogeneous reaction can be broken up into three steps:

Adsorption and dissociation of the reactants from the gas phase.

Chemical rearrangement on the surface.

Evolution of products.

It is our intention to examine each of these steps for a simple polyatomic gas through the coordinated application of a variety of experimental techniques.

One of the big questions in studies of this type is chemical purity of the reagents. As a first step in this area, field-emission studies will be undertaken, as these afford a direct indication of the

cleanliness of both the metal surface and the colliding gas. Work has begun on a general-purpose, field-emission microscope system for exploratory surface examinations.

Once this effort is under way, studies of dissociation at a crystal surface will begin. These will draw heavily upon our work on adsorption layers on single-crystal planes, using the field-emission microscope. In addition a mass spectrometer in direct line of sight with a crystal will be employed to sample the reaction products. This technique should provide information both on the initial act of decomposition and also on the subsequent chemical rearrangements occurring on the surface.

A separate stainless-steel, ultrahigh-vacuum system to accomodate a spectrometer has been started. With delivery of a quadrupole mass analyzer expected during the next few weeks, work will commence on identifying the products of chemical transformations occurring on metal surfaces.

1.4.3. Molecular-Beam and LEED Studies

Part of the effort already outlined will be carried out on macroscopic crystals, prepared by ordinary metallurgical techniques. Only in this way will sufficient surface area be available to follow the course of chemical changes. Although this will involve a sacrifice in the perfection of the surface, compared to that achievable in a field-emission microscope, it will establish a link to systems actually realizable in practice that could conceivably be utilized in future technology. In order to characterize these crystal surfaces we expect to resort to low-energy electron diffraction as well as to electron-stimulated Auger ejection [20]. A system permitting studies of this kind has been acquired, and we are now involved in instrumenting this.

It is planned to utilize this equipment for molecular beam studies as well. The aim is to examine the efficiency of energy transfer between a colliding atom or molecule and the lattice. It is this step that will largely dictate the subsequent course of the interaction. If sufficient energy is taken up by the lattice, the incident particle will be captured by it; otherwise, it will return to the gas phase. The details of the energy-transfer process will also affect the behavior of the entities that are captured. They may be left in a highly excited state, of enhanced chemical reactivity. Of special interest, therefore, will be studies on

chemically reactive systems for which energy transfer has not been examined in the past. Effort, however, will only get underway once diffraction and mass-spectrometric work has been properly instrumented.

G. Ehrlich
R. S. Polizzotti
R. G. Rowe
C. N. Stewart, Jr.

1.5. Atomic Exploration of Crystal Surfaces[†]

A variety of techniques are presently available to determine the behavior of adsorbed layers on crystals. As a consequence, we are on the way toward achieving accurate information about adatoms, from the atomic level on up. The state of understanding of crystal surfaces themselves is considerably less advanced. Despite the importance of this field, until a few years ago little information was available concerning the structure of crystal surfaces or the properties of individual atoms held on their parent lattice. This deficiency arose largely because of the extreme difficulties of carrying out experiments. The invention of the field-ion microscope in the 1950's has changed the situation dramatically [21].

This instrument is essentially a field-emission microscope operated backwards. With the screen negative, helium at a pressure of approximately 0.1 N/m^2 is introduced into the system. Close to the surface, electrons tunnel from the helium into the metal, leaving behind positively charged ions which are accelerated toward the screen. An image of the surface with a resolution somewhat under a nanometer can easily be obtained. With the field-ion microscope it is now possible to visualize individual metal adatoms as well as to resolve the detailed atomic arrangement of many crystal planes. It is our aim to exploit this technique in exploring the behavior of single metal atoms.

Initially the main theme emphasized in this effort will be the

[†]Supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy and U.S. Air Force) under Contract Number DAAB-07-67-C-0199.

surface migration of adatoms. Our work will be aimed at identifying the important physical parameters that affect the individual steps in diffusion--adatom motion over perfect flats, interaction with defects as well as other adatoms, and incorporation into lattice steps.

A start has been made at implementing these aims. A stainless-steel ultrahigh-vacuum system for field-ion microscopy is under construction. The microscope tube itself is now undergoing modifications, and both should be operational this summer. In the longer run, it is hoped to supplement this with an atom probe [22]--Müller's most recent invention that allows a measurement of the charge-to-mass ratio of a single atom, the location of which has been defined by field-ion microscopy.

G. Ehrlich
G. Ayrault
R. S. Chambers

1.6. References

1. Progress Report for March-August 1967; Progress Report for September 1967-February 1968; Progress Report for March-August 1968, Coordinated Science Laboratory, University of Illinois.
2. M. Nishijima and F. M. Propst, "Electron Bombardment-Induced-Ion Emission from Surfaces," Bull. Am. Phys. Soc. 2, 189 (1969).
3. F. M. Propst and M. Nishijima, "Electron Impact Desorption of Ions and Neutrals from Polycrystalline Tungsten," 29th Physical Electronics Conference, March 17, 1969, Yale University, New Haven, Conn.
4. M. Nishijima, "Electron Impact Desorption of Ions and Neutrals from Polycrystalline Tungsten," Report R-410, April 1969, Coordinated Science Laboratory, University of Illinois.
5. W. M. DuMond, Ann. Phys. 2, 283 (1957).
6. P. A. Redhead, Supplemento al Nuovo Cimento, Serie 1, 5, 586 (1967).
7. T. Madey and J. T. Yates, Jr., Surface Science 11, 327 (1968).
8. J. W. Coburn, Surface Science 11, 61 (1968).
9. P. A. Redhead, Can. J. Phys. 42, 886 (1964).
10. W. Fite and R. T. Brackmann, Phys. Rev. 113, 815 (1959).
11. D. Menzel, Ber. Bunsenges, Physik. Chem. 72, 591 (1968).
12. D. Menzel and R. Gomer, J. Chem. Phys. 41, 3311 (1964).

13. J. T. Yates, Jr., T. E. Madey and J. K. Payn, Supplemento al Nuovo Cimento, Serie 1, 5, 556 (1967).

14. Progress Report for March-August 1968, Coordinated Science Laboratory, University of Illinois.

15. Progress Report for March-August 1967, Coordinated Science Laboratory, University of Illinois.

16. E. J. Scheibner and L. N. Tharp, Surface Science 8, 247 (1967).

17. C. J. Powell, J. L. Robins, and J. B. Swan, Phys. Rev. 110, 657 (1958).

18. G. A. Harrower, Phys. Rev. 102, 340 (1956).

19. R. H. Good and E. W. Müller, Handbuch der Physik XXI, 176 (1956).

20. L. A. Harris, J. Appl. Phys. 39, 1419 (1968).

21. E. W. Müller, Science 149, 591 (1965).

22. E. W. Müller, J. A. Panitz, and S. B. McLane, Rev. Sci. Instr. 39, 83 (1968).

R. N. Peacock
G. E. Anner
R. Birtcher
W. P. Bleha
J. W. Culton
S. W. Depp
R. L. Jimenez
O. G. Riley
J. E. Robinson
B. Tao

2.1. Infrared Converter[†]

The development of a thin-film, solid-state device that will directly convert infrared radiation into visible light is the objective of the present study. As conceived, the device would consist of a sandwich of germanium, an insulating layer of less than 10nm thickness, and a layer of cadmium sulfide. Holes generated by the absorption of infrared light in the germanium would, under proper conditions of bias, tunnel into the CdS and generate visible light upon recombination with electrons. Although simple in principle, the development of such a device is by no means trivial. Considerable effort has been expended in determining the parameters that affect the properties of thin films of cadmium sulfide. As will be seen in this report, considerable success has been achieved in making films whose luminescent properties closely resemble those of single crystals. A major effort has also been necessary to determine the circumstances under which tunneling can be said to be positively responsible for electroluminescence in the so-called Jaklevic-type diodes. As will be seen below, some considerable uncertainty remains in establishing this. Work will continue on each of these areas during the coming quarter. The importance of the insulating barrier to tunneling is of such significance that a separate program in the determination of the properties of tunneling barriers in semiconductors is being initiated. These three efforts are expected to provide the basic information that is necessary for developing a device of the type described above.

2.1.1. Cadmium Sulfide films, Experimental Results for Resistivity and Mobility Measurements

Work has continued on the characterization of the electrical and luminescent properties of vacuum-evaporated, polycrystalline-CdS films.

[†]Supported principally by the Advanced Research Projects Agency under U.S. Army contract DAAK-02-67-C0546 and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199); and in part by the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

The substrates used were fused quartz and soda-lime glass. The evaporation technique used has been reported previously. The samples were masked in typical van der Pauw shape. The temperature of the sample holder could be varied from 77 K to above room temperature, but resistivity measurements were not practical over the entire range. The results of these measurements will be presented in this section.

The temperature dependence of the dark resistivity of various samples (no post-evaporation treatment) is shown in Fig. 2.1. The straight lines observed in the plots of log p vs $10^3/T$ at temperatures above 250 K, are indicative of an exponential behavior of the resistivity of the samples, $p = p_o \exp E_p/kT$. It is possible to obtain the activation energy E_p for the straight-line segment of the plot of p vs $10^3/T$ from the following relationship:

$$E_p = \frac{k\ln(p_1/p_2)}{1/T_1 - 1/T_2}$$

where k is the Boltzmann constant, p_1 and p_2 are the resistivities at two points on the straight line, and T_1 and T_2 are the absolute temperatures corresponding to p_1 and p_2 respectively.

The activation energies for the different samples were calculated and are also shown on Fig. 2.1. At low temperatures, the resistivities of the films increased by more than two orders of magnitude. This causes the surface space-charge layer to play a larger role in the conductivity of the film, thereby explaining the unsystematic behavior of the various films in the low temperature range. A similar behavior in CdS films was reported in the collection of data by M. Neuberger [1].

Earlier in the experiment, high-resistance samples prepared from high-purity CdS, obtained from Eagle-Picher, were measured. They showed room temperature resistivities of the order of 10kΩm with activation energies of about 0.40eV.

Except for very-high-resistance samples, the noise produced by the sample was not large enough to have introduced any sizable error. The thermoelectric effects due to temperature gradients along the sample were eliminated by reversing the direction of the current flow and averaging the resultant voltage with that obtained when the current was in the opposite direction. In addition to the error introduced by the finite

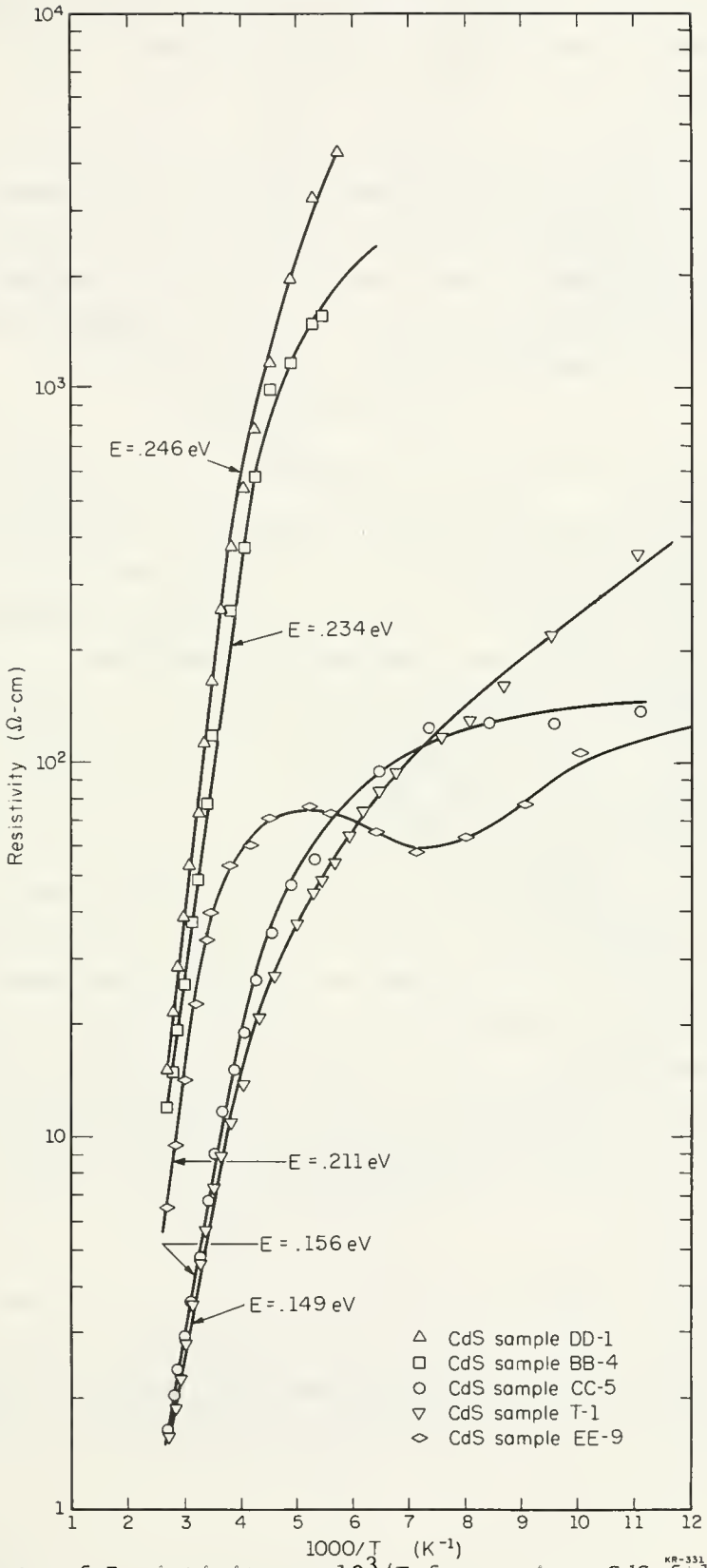


Fig. 2.1. Plots of Resistivity vs $10^3/T$ for various CdS films ($10^2\Omega$ cm = 1Ωm).

size of the contacts, other sources of error were due to temperature instability and to inaccuracies in sample thickness. These errors could total about $\pm$ 10%.

The temperature dependence of the "dark" Hall mobility for three of the films studied is shown in Fig. 2.2. The small values encountered for the mobility of the samples made it very hard to obtain accurate measurements. This was the main reason for restricting the temperature range over which the Hall mobility was observed. Again, the plots showed indication of an exponential form for the mobility,

$$\mu = \mu_o \exp(-E_\mu/kT)$$

Straight lines were fitted to the data by eye; then the mobility activation energies E_μ were obtained. The exponential variation of mobility was attributed by Shallcross [2] to intercrystalline barriers or imperfections within crystallites. The values of mobility observed were about two orders of magnitude lower than the reported single-crystal mobilities [3].

The mobility measurements were obtained by applying a constant magnetic field of 0.395T perpendicular to the film surface. The change in the van der Pauw resistance, $R_{BD,AC}$, was obtained from the change in voltage arising upon the application of the magnetic field while the current through the sample was kept constant. In the case of the CdS films, these changes were in the microvolt range. The accuracy in the measurement of these small voltages was limited by the noise produced by the sample. In order to minimize the error introduced by the noise, ten or more readings of $\Delta R_{BD,AC}$ were obtained and averaged. If $\Delta R_{BD,AC}$ was normally distributed, the probable error in the measurement of each point was about $\pm$ 5%. This error, together with the errors due to the finite size of the contacts, film thickness, and temperature instability, could total about $\pm$ 25%.

The room temperature Hall mobility μ_H, its activation energy E_μ, and the carrier mobility in the crystallite grains μ_o were the quantities determined. Table 2.1 shows a comparison of literature values for CdS thin films and typical values obtained in this experiment.

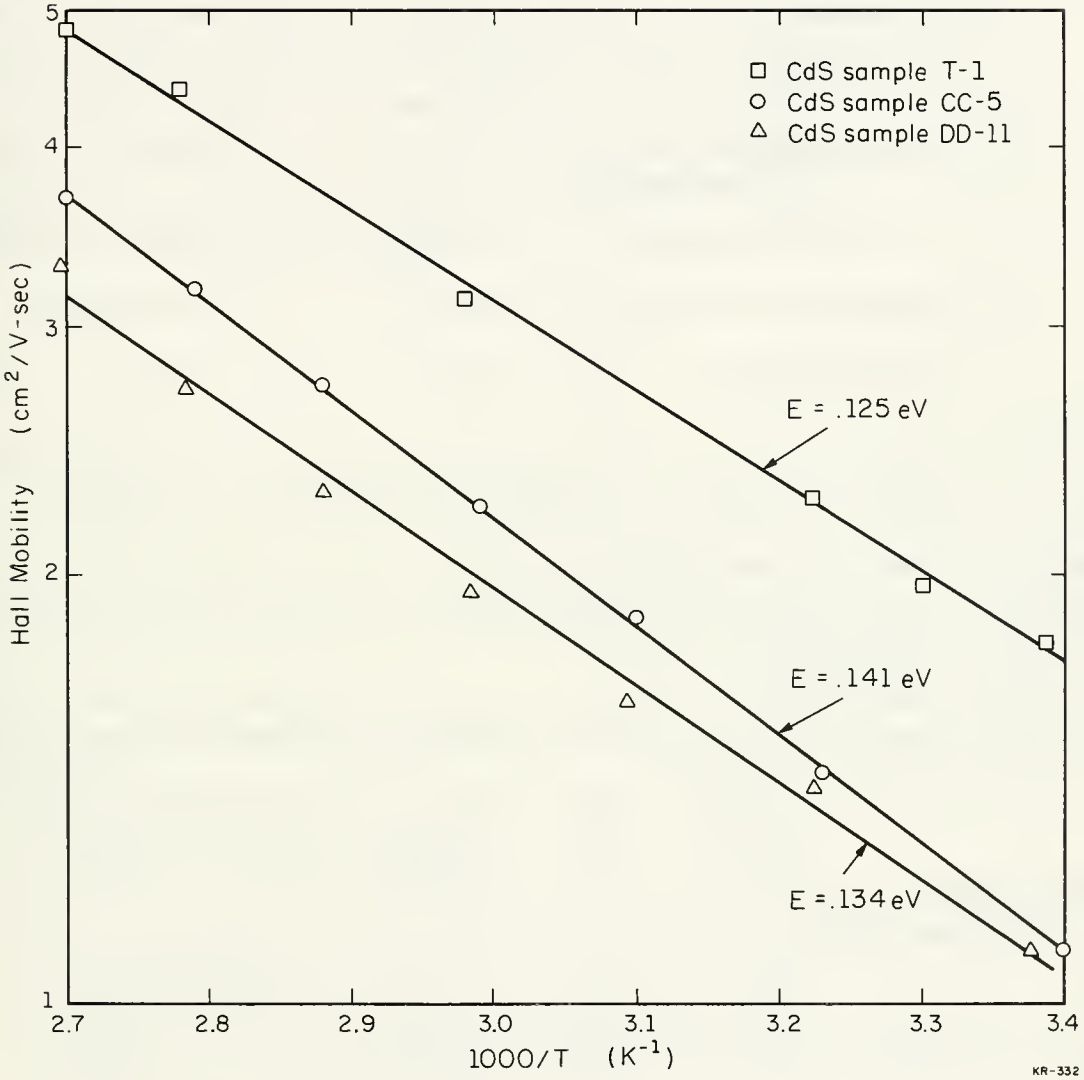


Fig. 2.2. Plots of Hall mobility vs Inverse Temperature for various CdS films (10^4 m^2/V s = 1 m^2/V s).

Table 2.1. Comparison of Mobility Measurements with Reported Values

Source	μ_H at 300K (m^2/Vs)		μ (m^2/Vs)		E_μ (eV)
H. Berger [4]	5	$\times 10^{-4}$	1	$\times 10^{-2}$	0.12
R. G. Makarious [5]	12	$\times 10^{-4}$	2	$\times 10^{-2}$	0.12
I. A. Karpovich and B. W. Zvonkov [6]	0.3	$\times 10^{-4}$	1	$\times 10^{-2}$	0.15
F. V. Shallcross [7]	5	$\times 10^{-4}$	1	$\times 10^{-2}$	0.074
Nguyen Duy Thuoc [8]	2	$\times 10^{-4}$	$T^{3/2}$ dependence		
This experiment	1	$\times 10^{-4}$	2	$\times 10^{-2}$	0.13

From the values of resistivity and mobility, it is possible to obtain the carrier concentration, and thereby the Hall coefficient. Table 2.2 shows the room temperature values of p, μ, and n for the various samples. The exponential variations of E_p and E_μ above 250 K are also included in this table.

Table 2.2. Resistivity, Hall Mobility and Carrier Concentrations of CdS Films

Sample	Thickness (μm)	p at 300K (Ωm)	E_p (eV)	μ_H at 300K (m^2/Vs)	E_μ (eV)	n at 300K (m^{-3})
T1	1.4	0.048	0.15	1.9×10^{-4}	0.13	6.8×10^{23}
CC5	1.3	0.055	0.16	1.1×10^{-4}	0.14	1.0×10^{23}
EE9	1.6	0.33	0.21	1.1×10^{-4}	0.13	1.7×10^{23}
BB4*	1.3	0.70	0.23	...	...	...
DD11	1.4	0.90	0.25	1.0×10^{-4}	0.13	6.9×10^{23}

*Sample too noisy to be able to determine the mobility with any reasonable accuracy.

Examination of the data presented in Table 2.2 reveals a systematic variation of the resistivity p and the activation energy E_p for the samples tabulated. Upon further investigation, it was found that if the logarithm of $\sigma_o = 1/p_o$ for each sample is plotted against its respective conductivity activation energy, these points fall on a straight line. This phenomenon was previously reported for various compound semiconductors by W. Meyer and H. Neldel [8], and more recently on annealed CdS films by H. Berger [4]. A theory explaining this phenomenon was developed by I. Broser and R. Broser-Warminsky [9]. They showed that if the equation defining the straight line is written

$$E_p = C_1 + C_2 \log \sigma_o,$$

where

$$\sigma = \sigma_o \exp(-E_p/kT)$$

then

$$E_p = N_A/N_d \tag{1}$$

and

$$-C_1/C_2 = \log(\epsilon \mu N_C) \tag{2}$$

in which σ_o is the conductivity as $T \to \infty$, ϵ is the electron charge, N_A is the acceptor concentration, N_D is the donor concentration, μ is the carrier mobility within the crystallites, and N_C is the effective density of states in conduction band. In order for the equations above to be valid, the following conditions must be met:

1) The semiconductor is extrinsic with a high degree of compensation.
2) The mobility is independent of the position of the Fermi level in the bandgap.
3) There exists a broad distribution of levels having exponential dependence on energy within the gap.

The plot of $\log\sigma_o$ versus E_p is shown in Fig. 2.3. The values of C_1 and C_2 have been calculated and are shown along with the curve.

Since the CdS film is a highly compensated semiconductor and the values of mobilities are fairly constant within the range studied, it can be concluded that the straight line observed is an indication that the donor levels show a broad distribution, having an exponential behavior with energy, rather than discrete levels, or a uniform distribution of levels over an energy range of the bandgap, and that the activation energy is dependent on the donor concentration.

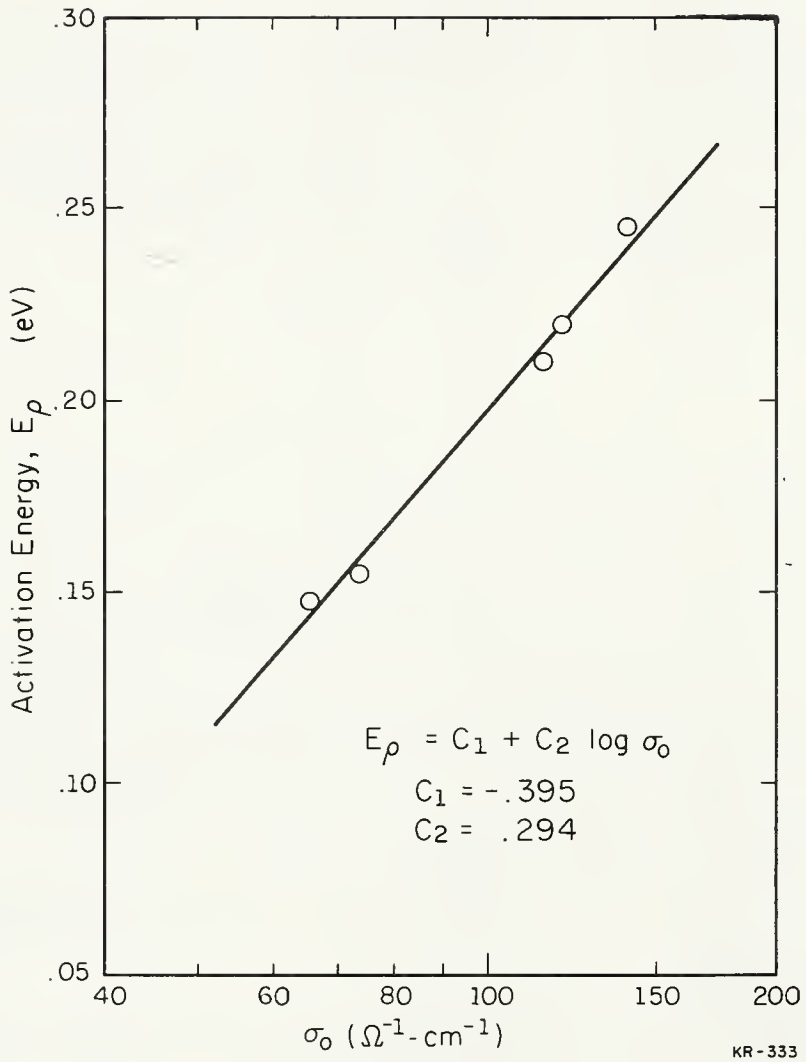


Fig. 2.3. Meyer-Neldel plot ($10^2\Omega$ cm = 1Ωm).

The value of 1.34 obtained for the ratio of the constants in Eq.(2) is comparable to 1.85 calculated for single-crystal CdS assuming $m_e^* = 0.2m_e$, $\mu = 2.0 \times 10^{-2}$ m^2/Vs and to 2.13 obtained by H. Berger for various CdS films annealed at different temperatures. Although the variations of resistivity activation energies were explained, it is still not clear what conditions in the deposition process of films determine the observed changes in the resistivity. The most likely variable would be variations in the impurity content in the source material and variations in the deposition rate.

The change in the resistivity and mobility of a sample, 1.3 μm thick, as a function of time of exposure to ambient was also examined. A plot of the room temperature values of p and μ versus time is shown in Fig.2.4. It can be observed that the resistivity increases very fast in the first few 8.64×10^4s (1 day) and then starts to level off. The total change of resistivity was estimated to be about 30% for the first 2.59×10^6s (1 month). The fact that the resistivity increases with time indicates the presence of acceptor surface states. It should also be evident that the conductivity of the sample is severely affected by the surface space-charge layer. This discussion does not imply that all of the films in this experiment behave in the same fashion, but it tries to point out the importance of the surface effects in the conduction processes of thin films. If the films were thicker, the influence of the surface would be expected to be lower. However, more severe changes in the resistivities of the films can be expected if the initial resistivity of the film were much higher. The mobility-time plot did not show any particular changes with time. Again the noise of the sample limited the accuracy of the measurements. To a first approximation, it can be stated that the mobility did not change with time.

Some films that had been subject to post-evaporation treatments were studied; three treatments were used.

One of the treatments was the baking of the CdS films, in saturated Cd vapor. This process was done in an evacuated closed tube. The pressure inside the tube was about 1.3 mN/m^2 (10^{-5} torr) at seal-off. High purity 6-9's Cd metal was used. The time of baking ranged between one and two hours while the temperature was varied between 450° and 600°C. The process was

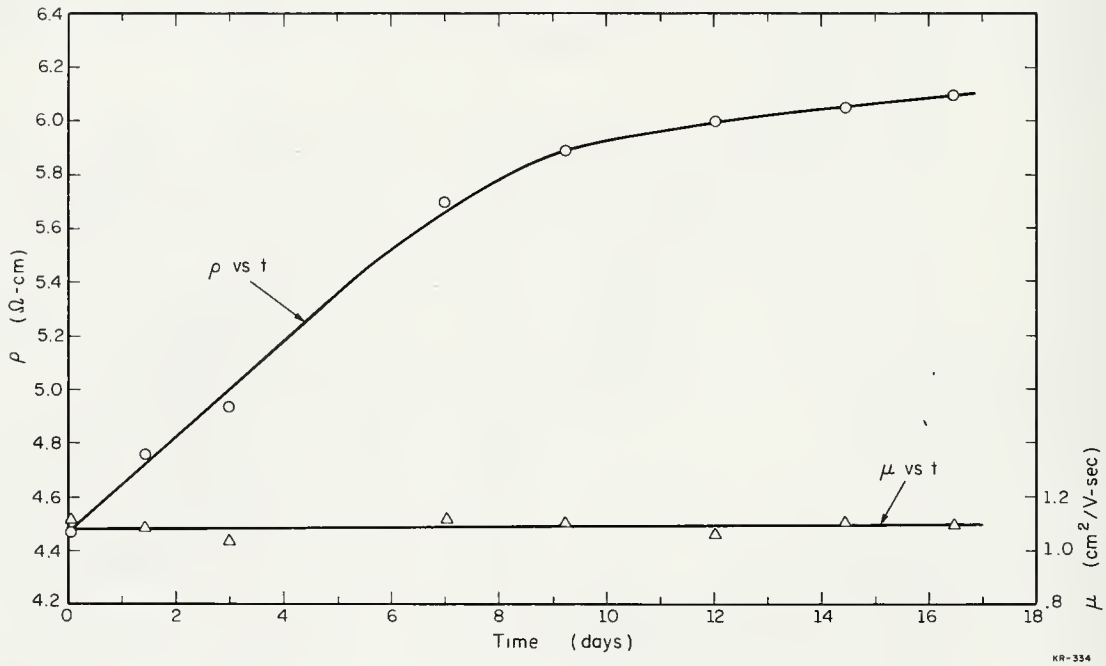


Fig. 2.4. Plots of the resistivity and the Hall Mobility of a CdS film vs the time of exposure to the ambient (1Ω cm = $10^{-2}\Omega$m).

terminated either by fast or slow cooling. When these processed films were placed on the measuring station, very large asymmetries were found in the van der Pauw measurements. Inspection of the samples by transmitted light with a microscope showed the existence of crazing in the films, on fused silica substrates. The microscope was a metallurgical type equipped with crossed polarizers.

The electrical asymmetry was so large that it did not permit measurements to be taken; still, it was possible to make some qualitative observations of the ranges of resistances encountered. It was found that samples which were quenched fast after processing had resistivities almost two orders of magnitude lower than those obtained in unprocessed films, while the samples that were allowed to cool slowly had higher resistivities than found in unprocessed samples.

Another treatment was a closed-tube vacuum anneal. This process is very similar to the one just mentioned above. The only difference is that no dopant is introduced into the tube. The temperatures, times, and tube pressure were allowed to be approximately the same as those used in the bake in saturated Cd vapor. Again, large asymmetries were observed in the measurements. They were also caused by crazing in the films. A qualitative measurement yielded values of resistivities many orders of magnitudes higher than the ones obtained for untreated samples.

The other treatment used was a bake in a gaseous atmosphere of a H_2S-H_2-HCl mixture. This was carried out at similar temperatures in an open-tube furnace with a gas flow. The measurements showed large asymmetries due to crazing. The resistivities of the films were observed to increase about three orders of magnitude. This increase was not as high as for the closed tube vacuum annealed samples.

The three post-evaporation processes had in common the fact that the samples were exposed to higher temperatures than those at which they were deposited. It has been observed that this situation would give rise to severe stresses on the films because of the differences in thermal expansion coefficients [10,11]. In this experiment, the stresses were so severe that crazing in the films resulted. From observations in the microscope, the crazing was found to be oriented; at this time no satisfactory explanation has been found, since the substrates are amorphous.

Figure 2.5 shows a microphotograph and an electron-scanning-microscope photograph of a film that was processed. Further investigation of this phenomena showed that the crazing did not occur for processed films that were deposited on glass. It was concluded, then, that the coefficient of expansion of the quartz substrates was too small compared to that of the film, and that the stresses in the sample due to this large difference were too severe for the sample.

R. Jimenez
W. Bleha

2.1.2. Photo Luminescence in CdS Films

Photoluminescent spectra have been recorded at both 13 K and 77 K for films of varying properties. Figure 2.6 (curve a) shows the spectrum at 77 K of a Cl-doped CdS film that has not been given any post treatment. The film shows green emission which doesn't show the phonon-replication peaks of single-crystal CdS. There is also a peak at 489.0 nm in the blue. Figure 2.6 (curve b) shows a similar film baked at 500°C in saturated Cd vapor. The green luminescence has diminished considerably in intensity while the blue peak at 489.0 nm has been enhanced. Figure 2.6 (curve c) shows a similar film baked at 600°C in vacuum in a closed ampule. The peak at 489.0 nm is reduced and the free excition peak at 487.5 nm has become apparent. The characteristic phonon structure of single crystal CdS has appeared in the green emission.

Previously only films baked in an H_2S-HCl flow gave the structured green emission. Now it has been found that vacuum baking in a closed tube has the same effect if the films contain the dopant required for edge luminescence.

Figure 2.7 shows spectra of a film baked in flow of H_2-H_2S-HCl at 650°C. The solid curve shows the spectrum at 77 K; the dashed curve shows the change with the sample at 13 K. One series results from recombination of free-to-bound carriers and the other from bound-to-bound recombination of the carriers.

Figure 2.8 is the spectra of similar films baked at different temperatures in a flow of H_2-H_2S-HCl. The dash-dot curve was taken on a

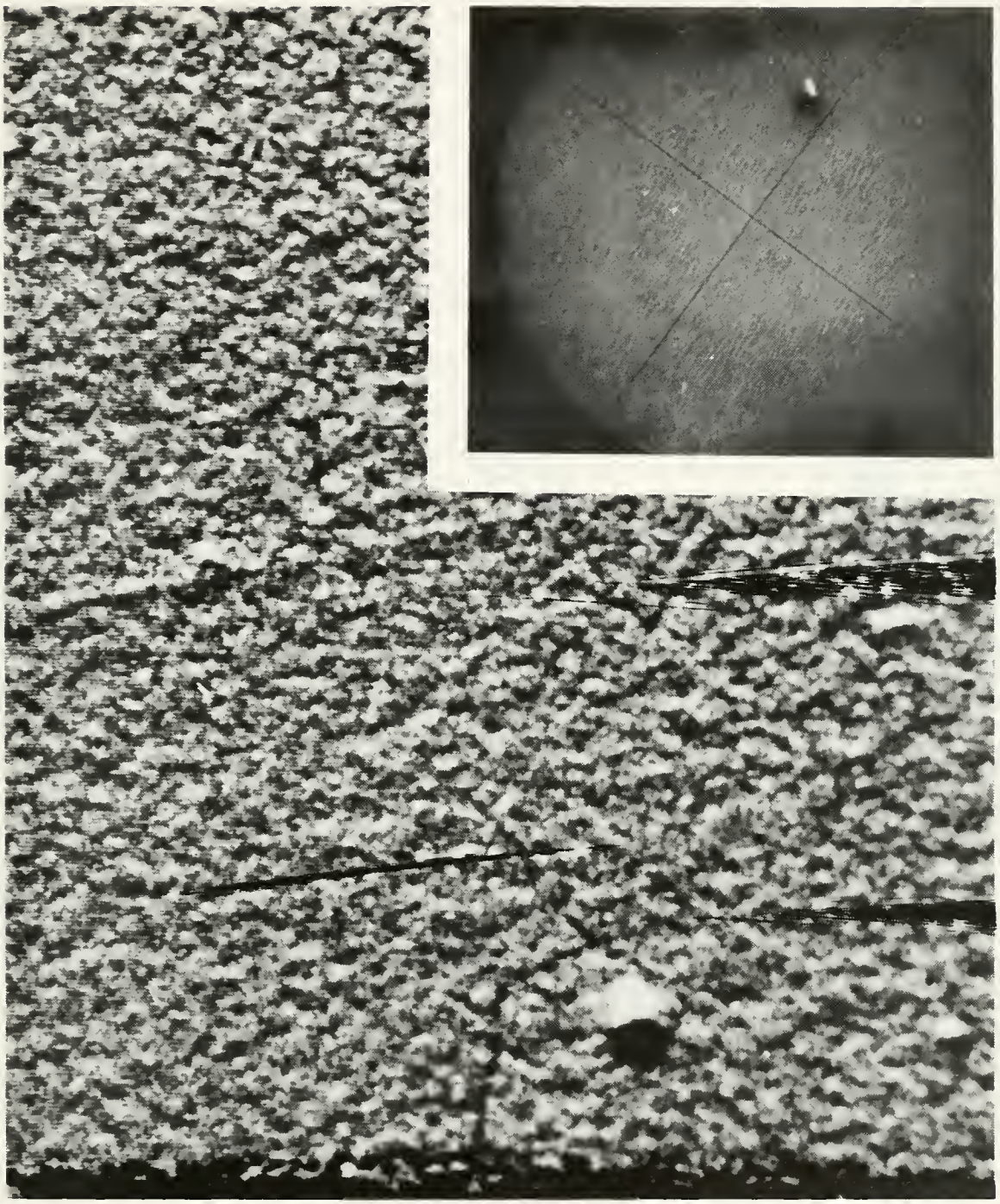


Fig. 2.5. Scanning Electron microphoto of crazing in a baked CdS film on a fused quartz substrate (magnification about 5000x). Inset: Optical microphoto of the same CdS film viewed by transmitted light at about 280x.

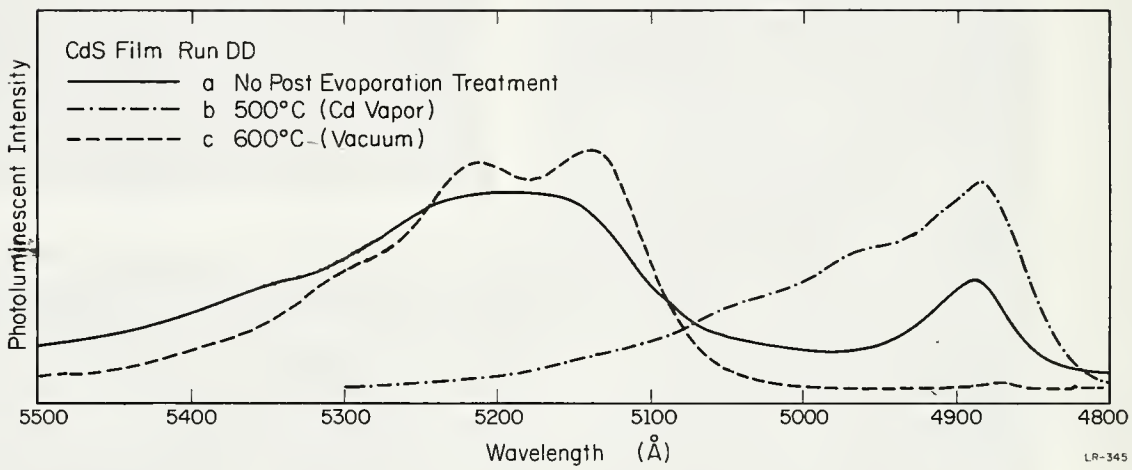


Fig. 2.6. Photoluminescent spectral changes of CdS films with different post evaporator processes. (10 Å = 1 nm)

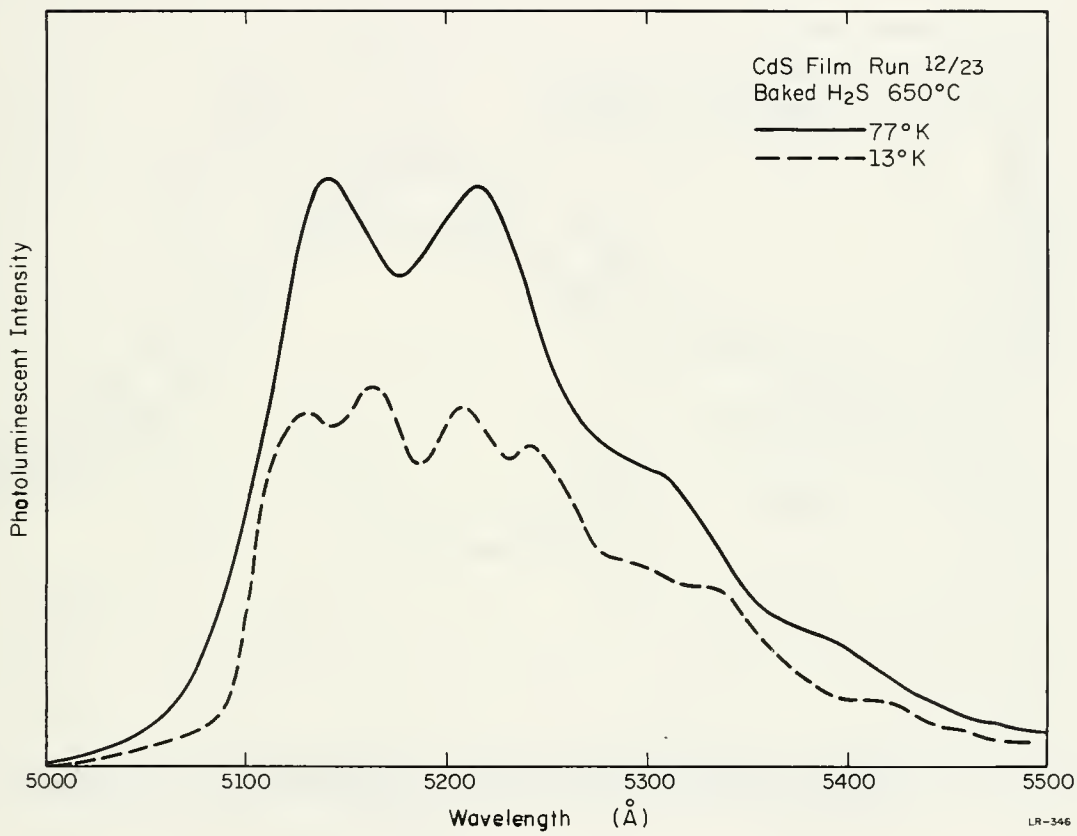


Fig. 2.7. Photoluminescent spectra of a H_2S baked CdS film at 77K and 13K. (10 Å = 1 nm)

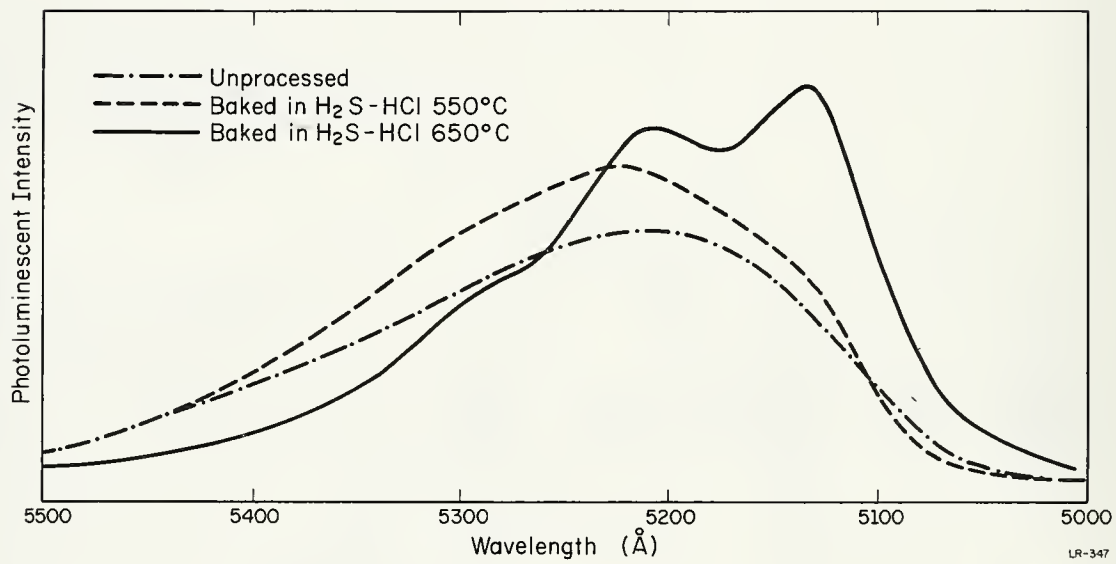


Fig. 2.8. Photoluminescent spectra of CdS films baked in H_2-H_2S-HCl at different temperatures. (10 Å = 1 nm)

film with no post treatment; the dashed curve on a film that was baked at 550°C; the solid curve is on a film that was baked at 650°C.

Figure 2.9 is the spectrum at 13 K and 77 K of an unprocessed film that has some structure at 77 K. The decrease in temperature to 13 K gives the characteristic edge emission of CdS. With the particular He dewar used, the exciton emission at 13 K was reabsorbed by the film, and was not recorded.

Time-decay measurements were made using a one-microsecond decay-time xenon flash tube to excite the luminescence. Figure 2.10 shows the time decay at 77 K for (a) a CdS film with no post treatment showing a broad photoluminescence peak at 522nm (77 K) (decay time < 1 μs). The measured decay is just that exhibited by the exciting flash tube and could be much shorter; (b) a CdS film baked at 500°C showing some structure decay time about 5 μs); (c) a CdS film showing edge-emission (decay time about 10 μs); (d) a CdS ultrahigh-purity single crystal (Eagle-Picher Company) (decay time about 20 μs). The decay time of the baked films approaches that of a single crystal while the unprocessed film is certainly shorter.

Measurements of photoluminescent intensity vs 1/T were taken to gain information as to the activation energy of the center (or centers) giving the observed luminescence. Figure 2.11 shows the temperature decay of the main peak of a CdS film showing edge emission. The observed activation is 0.17 eV for the baked film. Maeda [13] gives 0.165 ± 0.003 eV for the activation for insulating CdS single crystals so that the agreement is good with the baked films.

The resistivities of the doped films are on the order of 0.1Ωm while for the baked films they are greater than 10 kΩm. Table 2.3 summarizes these results.

From these results, it is seen that vacuum-evaporated CdS films that contain halogen impurities that have been baked in vacuum or H_2S at 500-600°C have luminescent properties that are very close to those of single crystals. By contrast, films that are not given a treatment after evaporation show structureless green emission and a peak in the blue. If the films are baked at temperatures below 500°C, they have properties intermediate between the untreated films and the single crystals.

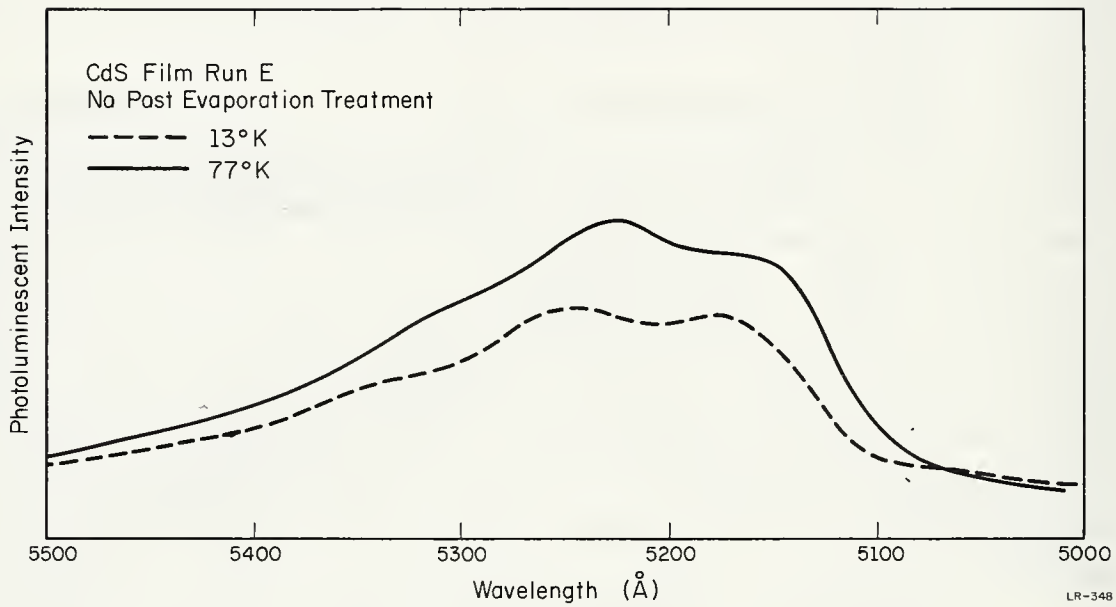
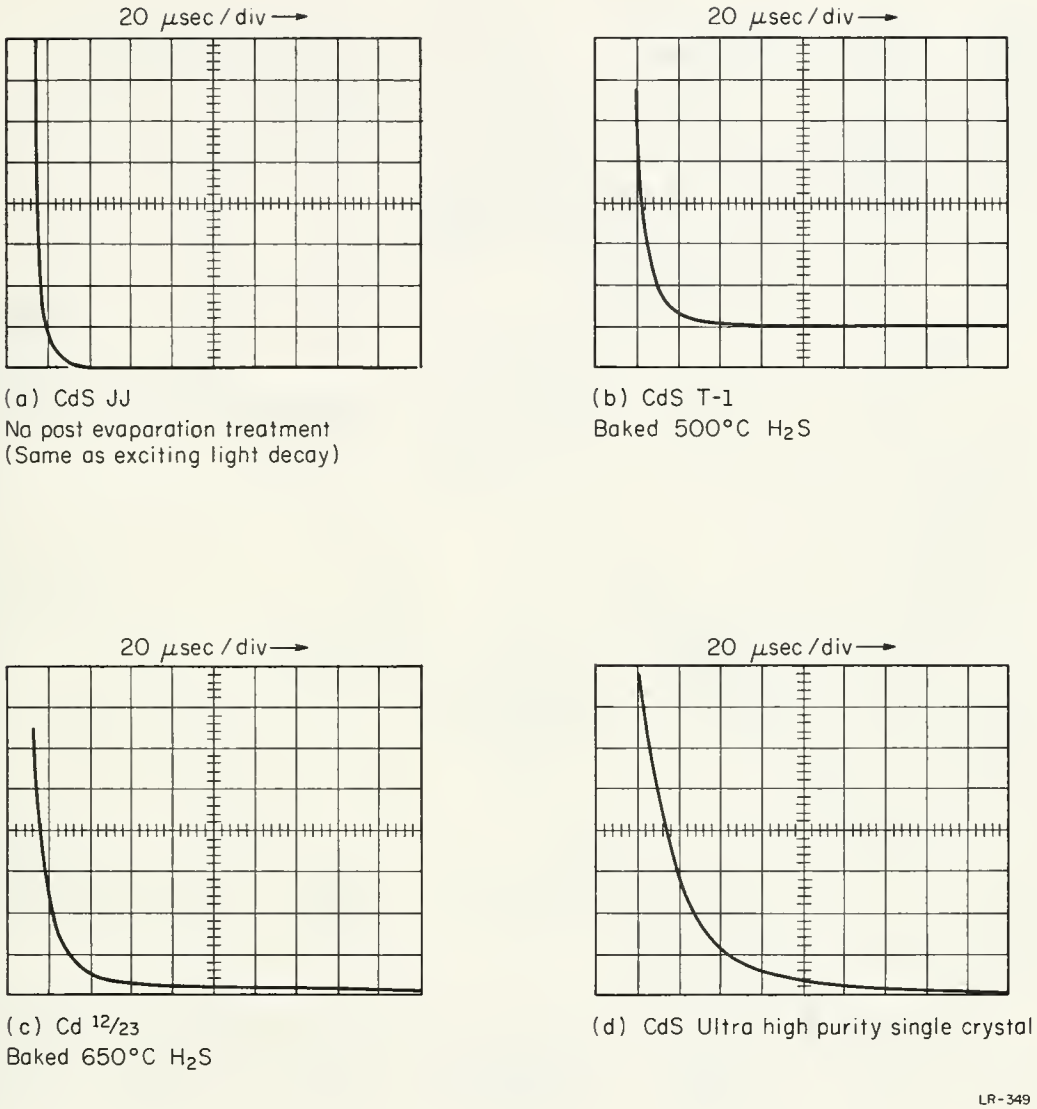


Fig. 2.9. Photoluminescent spectra at 13K and 77K of a CdS film having no post treatment. (10 Å = 1 nm)



(a) CdS JJ
No post evaporation treatment
(Same as exciting light decay)

(b) CdS T-1
Baked 500°C H_2S

(c) Cd $^{12}/_{23}$
Baked 650°C H_2S

(d) CdS Ultra high purity single crystal

Fig. 2.10. Time decay of green photoluminescence of CdS films and single crystal material.

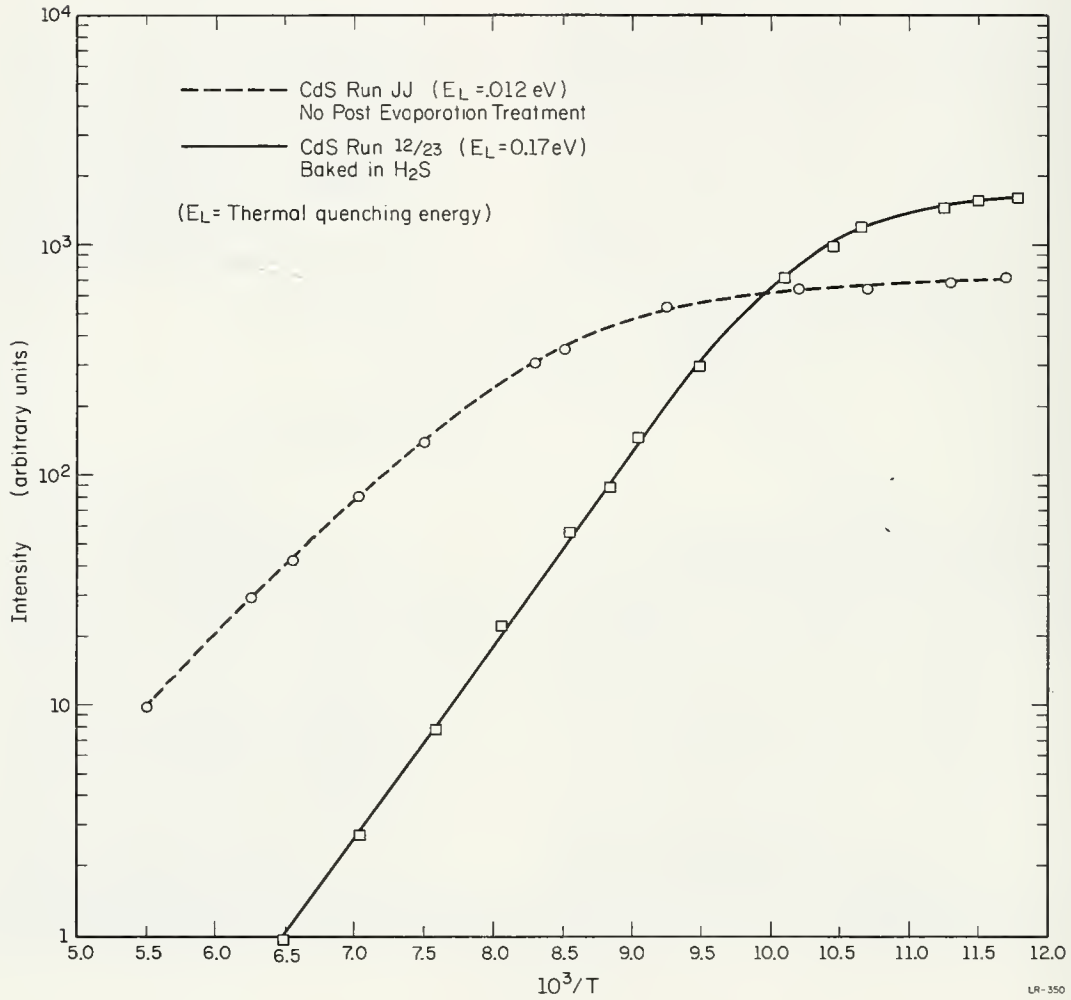


Fig. 2.11. Green photoluminescent intensity vs $10^3/T$ for CdS films.

Table 2.3. Summary of luminescent and electrical properties of CdS thin films.

	Resistivity 300K (Ωm)	Main Photoluminescence Peak 77 K (nm)	Characteristics of Green Photoluminescence	Photoluminescent Peaks in Blue (nm)	Green Photoluminescent Decay Time to 1/3 max (μs)	Green Photoluminescent Thermay Decay Activation Energy (eV)
Cl-doped CdS Films (no post evaporation treatment)	0.1-1	522.0	structureless or unpronounced structure	489 (broad)	< 1	_____
"Pure" CdS Films (no post evaporation treatment)	> 100	if luminescent 522.0	if luminescent unpronounced structure	if luminescent 489 (broad)	_____	_____
Cl-doped CdS Films (baked H_2S at T < 600°C)	0.1-10^3	522.0	unpronounced structure	very weak emission	5	_____
CdS Films (baked vacuum or H_2S^T > 600°C)	> 104	514.0	"edge" emission	487.5 free exciton	10	0.17
CdS Ultrahigh Purity Single Crystals	0.1-10^8	514.0	"edge" emission	487.5 free emission	20	0.165

Of particular interest are the relationships between CdS edge emission and the green photoluminescence found in unprocessed films and films baked at low temperatures. Possible explanations for the differences in luminescence are:

1. Recrystallization which results in smaller defect and imperfection concentrations when the films are baked. Also a decrease in the optical absorption in the green spectral region results when the films are baked. This absorption would mask structure, particularly at the shorter wave lengths. The tail of the absorption edge extends into the green emission region but a film having no structure at all had an increase in optical density from 522.0 nm to 514.0 nm of 0.06.
2. Baking causes an increase of centers which participate in the edge luminexcence and a decrease in the centers that give the unstructured emission.
3. Initial variations in the impurity concentration in the unprocessed films. With higher concentrations, the unstructured luminescence predominates; as the concentration decreases the edge emission increases in intensity.

The sharpness of observed structure is limited by the crystalline imperfection of the films.

Recrystallization of the films has been inferred by the exciton emission seen in the baked films [14]. A peak close to one of the CdS free-exciton peaks has been observed at 487.5nm in the baked films. Also the free excitons have been observed in optical absorption in baked films and not in unprocessed ones.

The scanning electron microscope has not shown any change in surface properties between baked and unbaked films. Observations are to be made by x-ray diffraction and electron microscopy in the future.

There is considerable evidence in the literature that impurity concentration affects the luminescent spectra. A peak develops at 522.0 nm with a shoulder at 514.0 nm (77 K) upon addition of impurity which becomes a near-gaussian-shaped peak with still further addition of impurity. This

has been observed by Van Doorn [15] with Ga, Bube [16] with I_2, Uchida [17] with both NaCl and Cd, Ibuki and Ohso [18] with Cd, Lehmann [19] with In, and Park and Litton [20] with Li. Except for Bube, Ibuki and Ohso, and Park and Litton, these measurements were made on CdS powders that were doped with different concentrations. Van Doorn and Uchida presented spectra at both 4 K and 77 K. At 4 K the typical edge emission structure was recovered from the 522.0 nm peak structure (Fig. 2.12). But this was not the case for the highest doping concentrations used. In this case, the temperature of measurement resulted in a negligible spectral shape change [15]. Figure 2.9 shows the same enhancement of edge luminescence in an unprocessed film with a temperature decrease from 77 K to 13 K.

Park and Litton [20] differ with Ibuki and Ohso [18] and Van Doorn [15] on the effect of Na doping but, since preparation methods differ, this is hard to interpret.

The 489.0nm peak in the blue was observed by Ibuki and Ohso [18] in Cd doped CdS crystals. (The temperature of the doping was 800°C.) They infer that the peak is due to a Cd interstitial, or a Cd interstitial associated with a trivalent positive ion with an energy level about 0.038 eV below the conduction band. The peak does not become sharp at 4 K as do exciton peaks, but this may be due to high impurity concentration. Ibuki and Ohso [18] observe a decrease in intensity of the green emission with baking in saturated Cd vapor at 600°C as do Handelman and Thomas [21]. When the CdS crystal is baked in Cd vapor at 800°C Ibuki and Ohso observe a broad green emission which persists to 300 K.

The results obtained on CdS films correlate with these observations on single crystals and powders. Figure 2.1 (curve b) shows the spectrum of a film baked in saturated Cd vapor. The green luminescence is reduced in intensity to a great degree. The total intensity is also reduced so that the blue peak at 489.0 nm is not enhanced significantly by the baking. Vacuum baking eliminates the 489.0 nm emission, enhances the green emission, and recovers the phonon-replication structure consistent with the results of Handelman and Thomas on single crystals. The enhancement of the edge luminescence is probably due to the increase in Cd vacancy concentration which is, to best analysis, part of the acceptor center [22]. The decrease in the blue emission could be due to an out diffusion of excess Cd and the halogen donor found in the evaporated films.

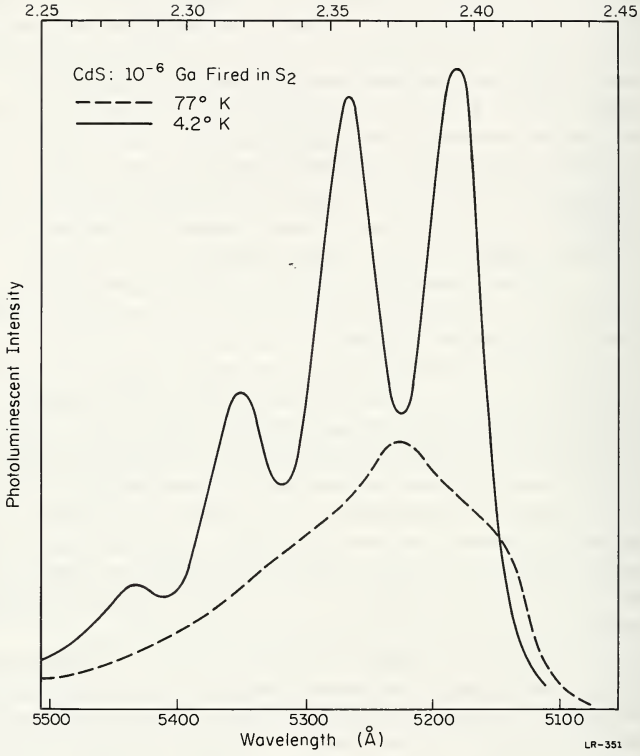


Fig. 2.12. Photoluminescent spectra of gallium-doped CdS powder at 77K and 4.2K.

Although Group-I impurities, such as Na and Li, may play a role in the green emission, the nature of this role is not clear. Van Doorn suggests that Na is the acceptor center involved in the edge emission and that a Cd vacancy complex is the acceptor involved in the unstructured emission. But this is not consistent with the work of Handelman and Thomas.

The present results suggest that the spectral variations found in unprocessed films are due to compeging luminescent processes. Since the two luminescent processes decay at different rates, it should be possible to separate them in time-decay measurements. Thus far, it has not been possible to get enough intensity at high resolution to record the spectra at delayed times. Observation at 4 K of both the blue and green luminescence should also prove to be useful. If the 522.0 nm luminescence is a bound-to-bound transition, there should be a shift in peak height with increase in intensity due to saturation of close paris. Again, a more intense exciting source must be used to check this.

W. Bleha

2.1.3. Tunnel Injection Luminescence

Data for electroluminescent turn-on voltages, light output vs voltage, and spectral output were obtained for several Jaklevic-type samples. The data to be reported were made using single crystal CdS, with the insulating film made by depositing a thin film of aluminum on the CdS and then thermally oxidizing it. Tunneling through the oxide film was confirmed in MIM structures, and degenerate CdS-Al-Al_2O_3-Pb sandwiches, by observation of the superconducting gap of thelead. This is not possible using an actual diode, however. Electroluminescence has also been studied in Jaklevic diodes using CdS thin films with results similar to those reported below for the crystals. The higher resistivity of the films, and of the contacts, reduce the value of this current-voltage data in making a basic interpretation, so the single crystal data has been selected here.

The samples were of the form CdS-Al-Al_2O_3-Au, where contact to the gold was made with either silver paint or a pressed-on indium ball. The light output was measured with a cooled photomultiplier (S-20 response). The threshold turn-on voltages ranged from 0.9-1.65V for forward bias (CdS neg) and from 1.2-3.2V for reverse bias. The intensities were roughly exponential with voltage for either bias direction, and some hysteresis was evident as shown in Fig. 2.13 and Fig. 2.14. The light output decreased with time, which was noted as a tailing-off of the high voltage end of the intensity-voltage curves. This was particularly evident in diodes in which the currents were greater than 100mA. The turn-on voltage also seemed to increase with the number of on-off cycles.

The electroluminescence spectrum was obtained for two of the brightest diodes, under forward and reverse bias. These are presented as Fig. 2.15 and Fig. 2.16. Although the resolution was limited by instrument noise, the spectra correspond roughly to typical edge luminescence spectrum for single crystal CdS, with the main peak at 523.0 nm. There was considerable output in the near infrared occurring as a broad peak centered around 985.0 nm.

The fact that all diodes for which electroluminescence was observed had turn-on voltages at considerably smaller biases than bandgap, and that some luminesced under reverse bias, tends to cast doubt on direct tunneling of holes as the injection mechanism in these samples. Work is presently in progress to determine which process or processes is responsible for the light emission.

G. Anner
D. Miller
B. Tao

2.1.4. Ge-Te Switching Units

Further work was done on the Ge-Te IR-sensitive switching diodes. Tests on units made by alloying indicate that the region of the heterojunction between the Te and Ge is the sensitive region. The tests were made with the IR beam from a HeNe (1.4 μm) laser scanned across the various

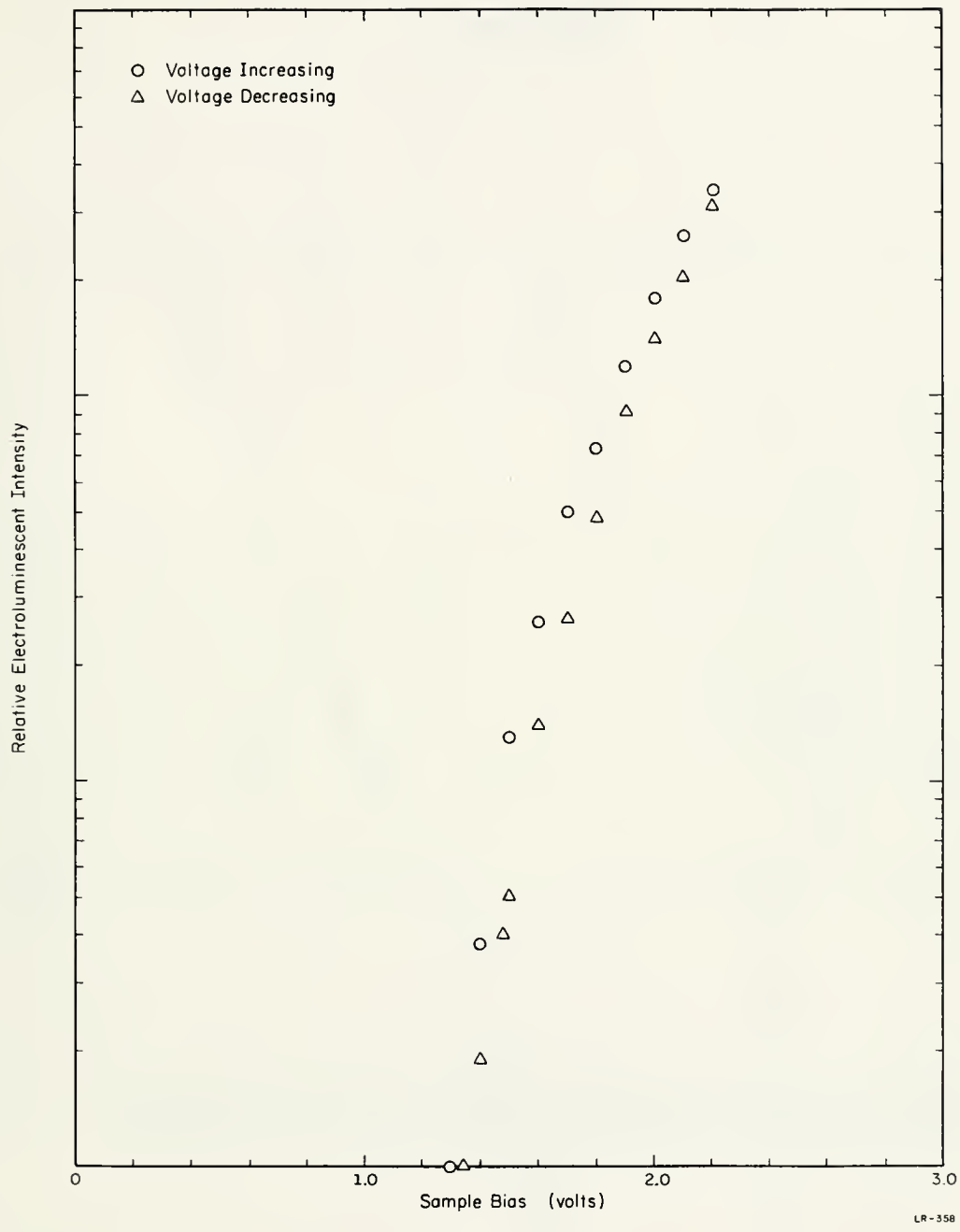


Fig. 2.13. Relative electroluminescent intensity of an In-CdS-Al-Al_2O_3-Au structure as a function of sample bias, CdS biased negative (forward direction).

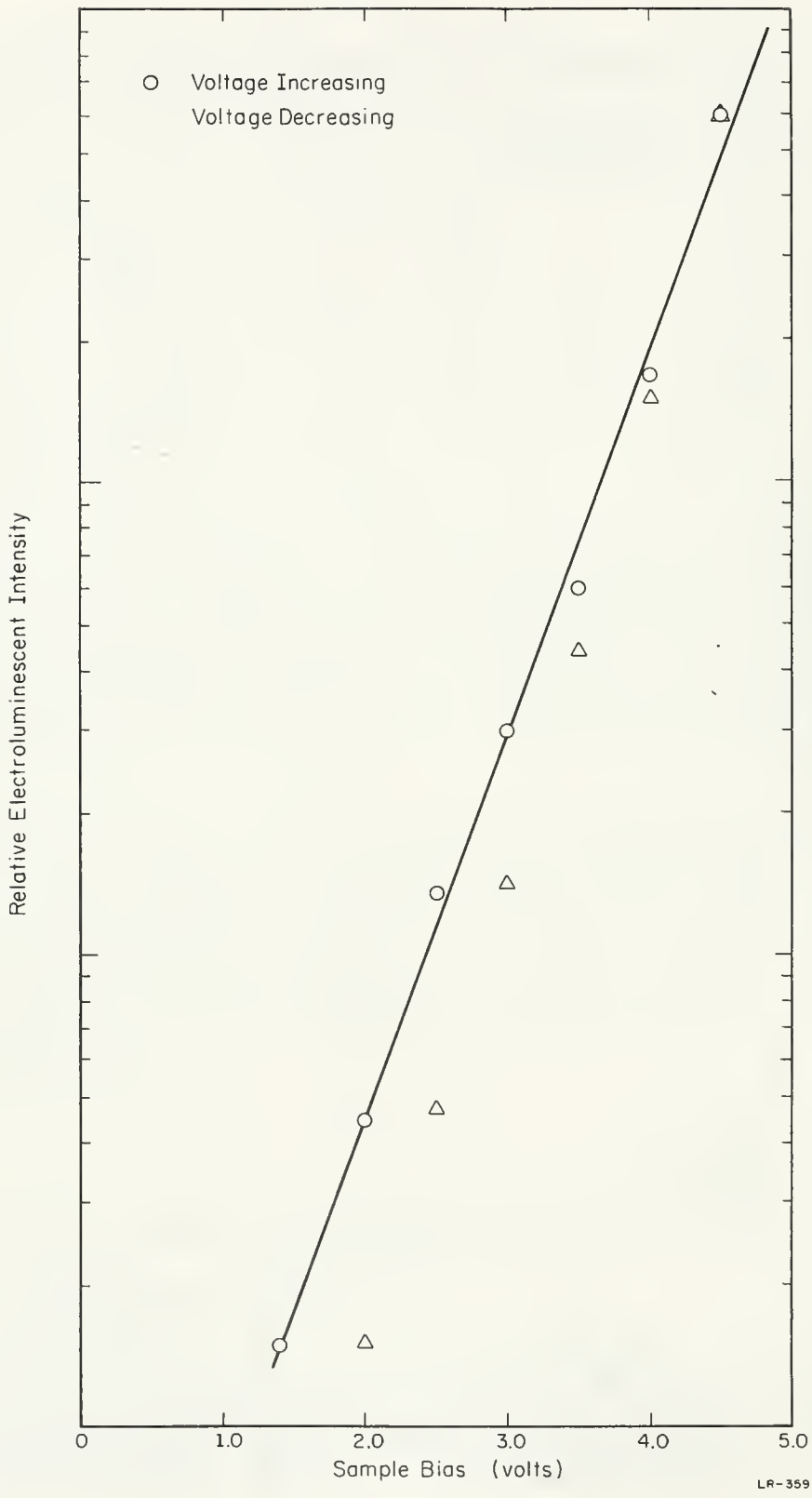


Fig. 2.14. Relative electroluminescent intensity of an In-CdS-Al-Al_2O_3-Au structure as a function of sample bias, CdS biased positive (reverse direction).

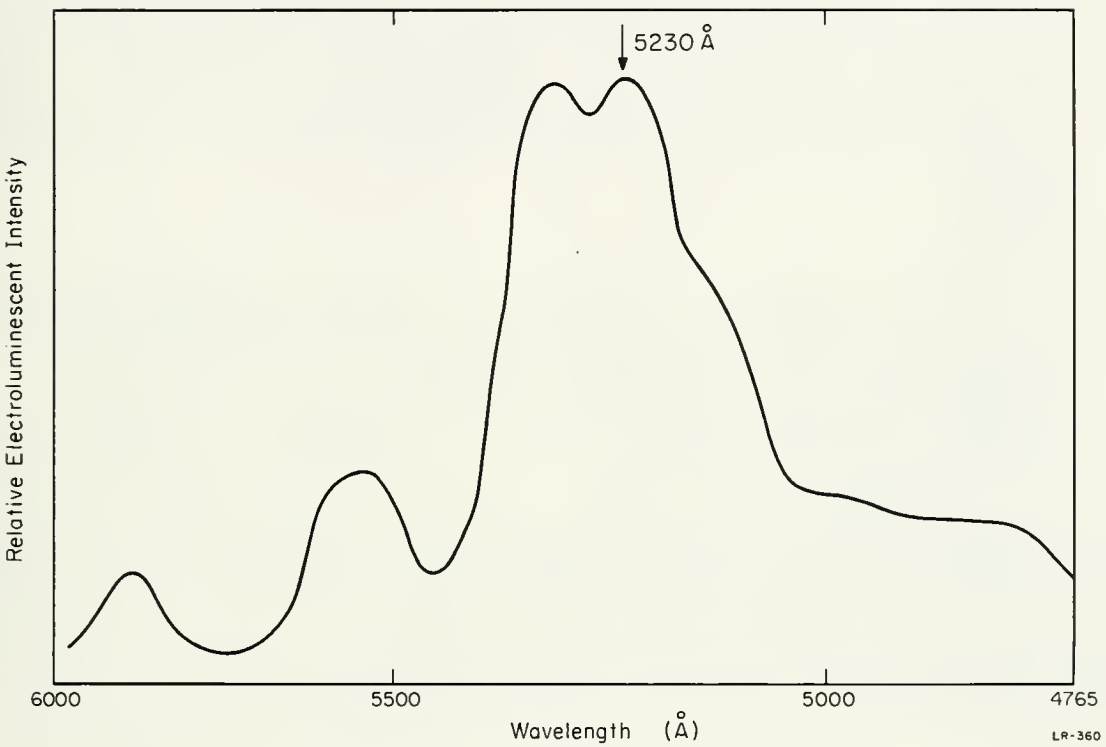


Fig. 2.15. Relative electroluminescent intensity of an In-CdS-Al-Al_2O_3-Au structure as a function of wavelength, CdS biased negative (forward direction). Instrument resolution approximately 6 nm.

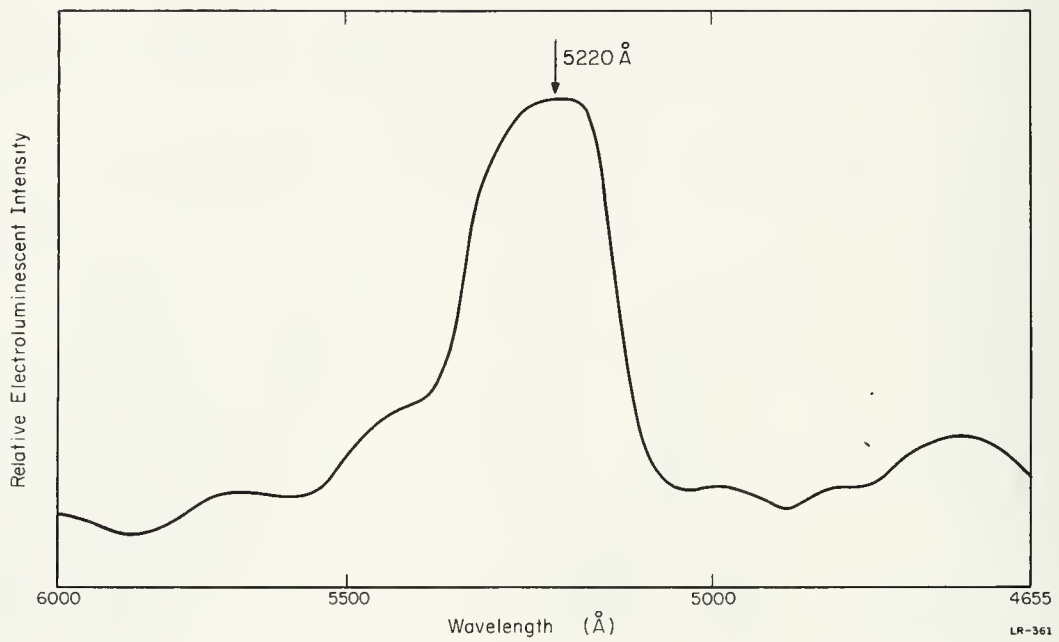


Fig. 2.16. Relative electroluminescent intensity of an In-CdS-Al-Al_2O_3-Au structure as a function of wavelength, CdS biased negative (forward direction). Instrument resolution approximately 6 nm.

regions of the diode. When the laser was operated at 623.8 nm no switching action occurred. It is suspected that surface effects prevent switching in this case.

Attempts were made to fabricate switching units by diffusion techniques. Te and 0.2-Ωm Ge were sealed off in a quartz ampoule at 0.4 mN/m^2 (3×10^{-6} torr) and heated at 800-850°C for one hour. One side of the die was removed and the unit assembled onto a header. It was found that unless the temperature was raised to roughly the peritectic temperature of the Ge-Te system during the process of attaching leads to the other upper surface, no switching was observed. This confirms earlier results obtained with the alloyed units.

G. Anner
B. Tao

2.2. Tunneling Barriers[†]

2.2.1. Parameters for Determining Barrier Characteristics

The major effort during this period was the investigation of various methods for determining the barrier heights in a metal-insulator-metal (MIM) tunnel junction. Work was confined to those methods which obtain the barrier heights through an analysis of the electrical characteristics of the junction. In order to determine the accuracy of these methods, the characteristics of a MIM junction were first calculated theoretically and then subjected to the various methods of analysis. If the calculations then yielded the barrier heights used in the original model, the method was deemed successful.

The first step in this effort was to calculate accurately the tunneling characteristics of a MIM junction for all possible barrier heights, thicknesses, and temperatures. A variety of approximate methods exist for this calculation but it was decided to obtain first the exact solution. The computer evaluation of the exact solution was extremely slow but did demonstrate that a much faster WKB method provides an adequate approximation. The WKB method was therefore used for all the calculations made.

The tunneling characteristics were then calculated for a variety of MIM junctions under the assumption of a uniform voltage drop across

[†]Supported principally by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199 and in part by the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

the insulating region (the trapezoidal barrier). Of the various methods for obtaining the barrier heights, the method of plotting the logarithmic derivative of the current vs bias voltage, d(lnJ)/dV vs V [24], seemed most successful. Peaks occur in this function at bias levels numerically equal to the barrier heights expressed in eV to an accuracy of less than 5%. Experimentally, this method is quite desirable, since it simply involves measuring the current and conductance as a function of voltage. Also, unlike the other methods, these measurements can be taken at a single temperature, which avoids any effect due to thermal cycling [25].

2.2.2. Trapped Charge in Insulator Barriers

The trapezoidal-barrier model for tunneling has been questioned when charge is present in the insulating film. This problem has now been examined theoretically. The model used was a metal-insulator-metal junction in which a positive charge was distributed uniformly in the barrier region. The effect of this charge for various combinations of barrier heights and thicknesses was investigated.

As a first step, the exact form of the barrier transmission probability $D(E_x)$ was found from a solution of Schroedinger's Equation in the junction region. Since a numerical evaluation of $D(E_x)$ required excessive computer time, it was decided to recalculate this quantity in the WKB approximation. The agreement of both methods was sufficiently good to allow the use of the WKB method throughout the remaining calculations.

It was found that trapped charge had the largest effect on the tunneling characteristics in the case of low barrier heights and large thicknesses. A symmetric barrier with the height of 0.5 eV and a thickness of 10nm was taken as standard and the effects of various charge densities, ρ calculated. It was found that for $\rho \leq 10^{24}\ m^{-3}$, a change in conductance of less than 20% was predicted. This value of ρ was well above the range considered physically reasonable. Hence the simple trapezoidal barrier remains a valid model as far as trapped charge is concerned. The result also insured the validity of the method discussed above.

S. W. Depp

2.3 Ionization-Gauge Studies in Low-Pressure Atmospheres[†]

Data has been taken to indicate the relative sensitivities of NASA flight-type gauges, commercial B-A gauges and Schuemann gauges in argon, helium, nitrogen, carbon monoxide, and oxygen. This data is presently under final analysis. An omegatron-like device was used as a secondary standard in relating pressures to ionization cross-section ratios for different gases. This device was used as both a partial- and a total-pressure instrument. The latter mode of operation was achieved by adding electron-beam defining apertures, by coaxially shielding the ion collector, and biasing it 30V negatively. The magnitude of the magnetic field was reduced from the normal 0.35T to 0.065T when employing it as a total pressure gauge. Previous progress reports have shown the relatively good comparisons obtained with published ionization cross section data and have illustrated effects in the device (and probably all omegatrons) which can result from improper bias arrangements. As a total pressure instrument, this device has an x-ray limit low enough for the purposes of these studies--13nN/m^2--a sensitivity for nitrogen of 0.19m^2/N and it functions well at emission currents of about 10 μA.

While the sensitivity data on the ionization gauges were being taken, magnetic fields became suspect as an explanation for discrepancies between gauge readings. These gauges had to be mounted on the vacuum system less than a meter distant from the flight-type magnetron gauge with its 0.1T magnet and the omegatron with its 0.35T magnet. To study this effect, a solenoid was constructed which could produce 10^{-2}T axial fields and yet contain a sealed off ionization gauge in its air core. As in studies [26,27] done at higher field strengths, the results indicate that magnetic fields parallel to the usual filament-to-grid electron path increase sensitivity while fields normal to this path decrease sensitivity. Figure 2.17 shows an example of this low-field data for the case of increased sensitivity in one model of B-A gauge. Similar results were found for the Schuemann gauge, conventional triode ionization gauge, and another design of B-A gauge. Since the magnetic fields about the gauges mounted on the test system were already reduced to less than 5×10^{-4}T, this effect is eliminated as a source of error in this study.

J. Culton

[†]Supported principally by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199 and in part by the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

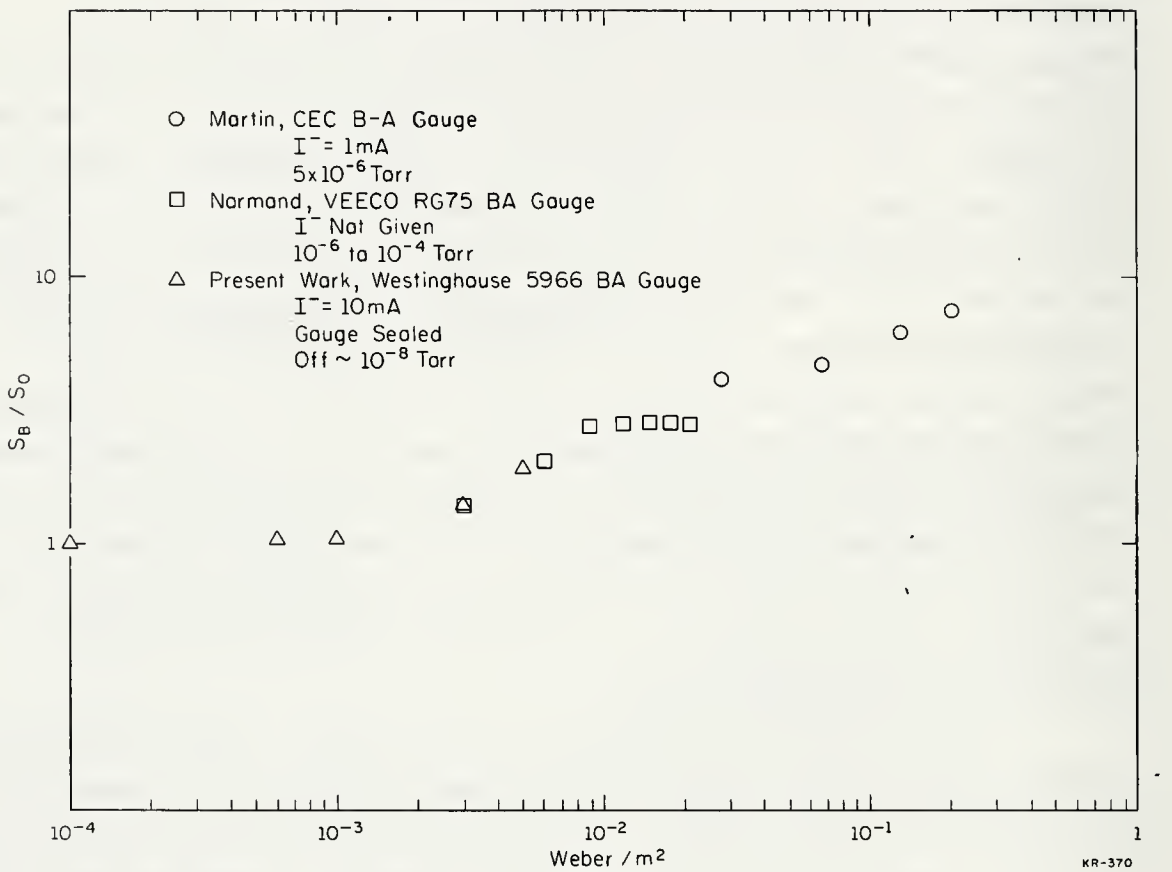


Fig. 2.17. Ionization gauge sensitivity in a magnetic field (normalized to the no-field condition) vs magnetic field.

2.4. Electrical-Discharge Machining[†]

Investigations were conducted during the past year using samples of brass, aluminum and tungsten which had been machined with the electrical-discharge machining process. It was hoped that some simplified method might be found for determining the approximate depth of damage which occurs during this type of machining. This damage, in single-crystal material, is important in preparing samples for devices or surface research. Some of the variables compared were:

1. moving wire vs solid electrode
2. relaxation circuitry vs pulse-type circuitry
3. copper surface vs brass surface vs tungsten surface.

The samples were viewed in the "as cut" condition using a scanning electron microscope (Cambridge Instruments "Stereoscan"). Figure 2.18 shows cuts made with both types of circuitry. The right portion is the pulse circuit, the left the relaxation circuit.

The results indicated that, when viewed at about 50X magnification, the appearance of the surface is dependent upon the type of material machined. It was also found that both the surface appearance and the depth of damage are dependent on the type of circuitry used, with the depth of damage inherently less with the pulse-type unit. Finally, use of the moving-wire electrode results in a more uniform and less damaged surface than the solid electrode, with a readily discernible difference in surface appearance. In some cases it has been possible to increase the cutting rate and decrease the damage.

The work is continuing in an attempt to obtain more definitive results. Combinations of voltage, amperage, capacitance, frequency, wire speed, and degree of preerosion of the wire electrode will be considered in addition to those parameters discussed above. Other materials, when possible single crystal, will be used.

G. Burr

[†]Supported principally by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199 and in part by the Physical Electronics Affiliate Program, College of Engineering, University of Illinois.

Fig. 2.18. An EDM cut as seen at about 400X with a scanning electron microscope. The right portion was cut with a pulse-circuit machine, the left with a relaxation-circuit type.

2.5 References

1. Cadmium Sulfide, ed. by M. Neuberger, Electronic Properties Information Center, DS-124/2E, Hughes Aircraft Co. (1967), pp. 42-43.

2. F. V. Shallcross, "Methods of Deposition and Physical Properties of Polycrystalline II-VI Films" R.C.A. Review 28, 569, 1967.

3. A. R. Moore and R. W. Smith, "Effect of Traps on Acoustoelectric Current Saturation in Cadmium Sulfide," Phys. Rev. 138, A1250 (1956).

4. H. Berger, "Uber das Ausheilen von Gitterfehlern frisch auf gedampfter CdS- Schichten (I)," Phys. Stat. Vol. 1, 739, (1961).

5. R. G. Maukarious, "Hall Mobility Measurements on CdS" Solid State Electron. 7, 702 (1964).

6. I. A. Karpovich and B. W. Zvonkov, "Hall Mobility of Electrons in Cadmium Sulfide and Selenide Films," Sov. Phys.-Solid State, 6, 2714 (1965).

7. Nguyen Duy Thuoc, "Properties Magnetogalvanics et Electriques des Couches de CdS Deposees Sous Different Vides," Compt. Rend. 263B, 826 (1966).

8. W. Meyer and H. Neldel, "Uber die Beziehungen Zwischen der EnergieKonstanten E und der MengeKonstaten a in der Leitwerts-Temperatur formel bei oxydischen Halbleitern," Z. Tech. Physik 18, 588 (1937).

9. I. Broser and R. Broser-Warminsky, "Statistisch-Kinetische Theorie der Lumineszenz und elektrischen Leitfahigkeit von Storstellen halbleiter," Ann. Phys. (N.Y.) 16, 361 (1955).

10. J. N. Zemel, "Transport Phenomena in Hetero Epitaxial Semiconductor Films," in The Use of Thin Films in Physical Investigations, ed. by J. C. Anderson (Academic Press, London and New York, 1966), p. 319.

11. R. W. Hoffman, "The Mechanical Properties of Thin Films," in The Use of Thin Films in Physical Investigations, ed. by J. C. Anderson (Academic Press, London and New York, 1966), p. 261.

12. W. Meyer and H. Neldel, Z. Tech. Physik 18, 588 (1937).

13. K. Maeda, J. Phys. Chem. Solids 26, 1419 (1965).

14. Quarterly Progress Report 1, Contract DAAK-05-67-C-0546. Coordinated Science Laboratory, p. 20.

15. C. Z. VanDoorn, J. Phys. Chem. Solids 29, 599 (1968).

16. R. H. Bube, RCA Review 20, 583 (1959).

17. I. Uchida, J. Phys. Soc, Japan 21, 645 (1966).

18. Ibuki and Ohso, J. Phys. Chem. Solids 27, 1753 (1966).

19. W. Lehmann, Solid-State Electronics 9, 1107 (1966).

20. Y. S. Park and C. W. Litton, Electronics, p. 104 (July 8, 1968).

21. E. T. Handelman and D. C. Thomas, J. Phys. Chem. Solids 26, 1261 (1965).

22. Thomas, Dingle, and Cuthbert, Proceedings of the International Conference of II-VI Semiconducting Compounds, Providence, 1967 edited by D. G. Thomas, (.W.A. Benjamin, Inc., New York, 1968), p. 863.

23. R. C. Jaklevic et al., Appl. Phys. Lett. 2, 1, 7 (January 1963).

24. K. Gundlach in Proc. Int. Symp. Thin Film Physics, Niedermayer and Mayer, eds. (Vandenbock and Ruprecht, Göttingen 1966) p. 696.

25. S. R. Pollack and C. E. Morris, Trans. AIME 233, 497 (1965).

26. G. D. Martin, Jr., in 1961 Trans. of the Am. Vac. Soc. (Pergamon Press, N.Y., 1961) p. 476.

27. C. E. Normand, in 1961 Trans. of the Am. Vac. Soc. (Pergamon Press, N. Y., 1961) p. 534.

M. Raether
H. Böhmer
D. Bollinger
J. Chang
K. Evans
D. Fenneman
R. Holden
E. A. Jackson
D. Kraybill
D. Stowe
M. Yamada

3.1. Influence of Collisions on the Growth Rate of the Electron Beam-Plasma Instability

Recently, we reported an investigation on the influence of collisions on the high-frequency instability in a beam-plasma system [1]. At that time, we used the spreading of the electron-beam energy-distribution function, which results from the interaction of the beam with the plasma, as a relative measure for the final amplitude of the instability. The advantage of this method is that measurements can be performed over a wide plasma-density range (10^{16} to 3×10^{19} m^{-3}). On the other hand, the change in the energy-distribution function which is observed after the beam has traversed the entire length of the plasma can be compared only qualitatively with the linear theory because nonlinear effects in the last part of the interaction region can influence the shape of the distribution function.

To check the linear theory quantitatively, we now have measured the linear growth rate of the beam-plasma instability in a plasma-density range from 10^{18} to 2×10^{19} m^{-3} and a wide beam-parameter range.

Let us recall the relevant equations for the interaction of a cold electron beam with a cold plasma. By the inclusion of a friction term in the equation of motion, one arrives at a dispersion relation

$$1 = \frac{\omega_p^2}{\omega(\omega+i\nu_c)} + \frac{\omega_B^2}{(\omega-kv_o)^2} , \qquad (1)$$

where ω_p and ω_B are the plasma and beam plasma frequencies respectively, ν_c is the collision frequency for momentum transfer, and v_o is the beam velocity. This dispersion relation has been discussed by Bludman, Watson, and Rosenbluth [2].

For the most unstable mode, one may write

$$\omega/\omega_p = kv_o/\omega_p+\varepsilon ,$$

†Supported principally by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199 and in part by the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

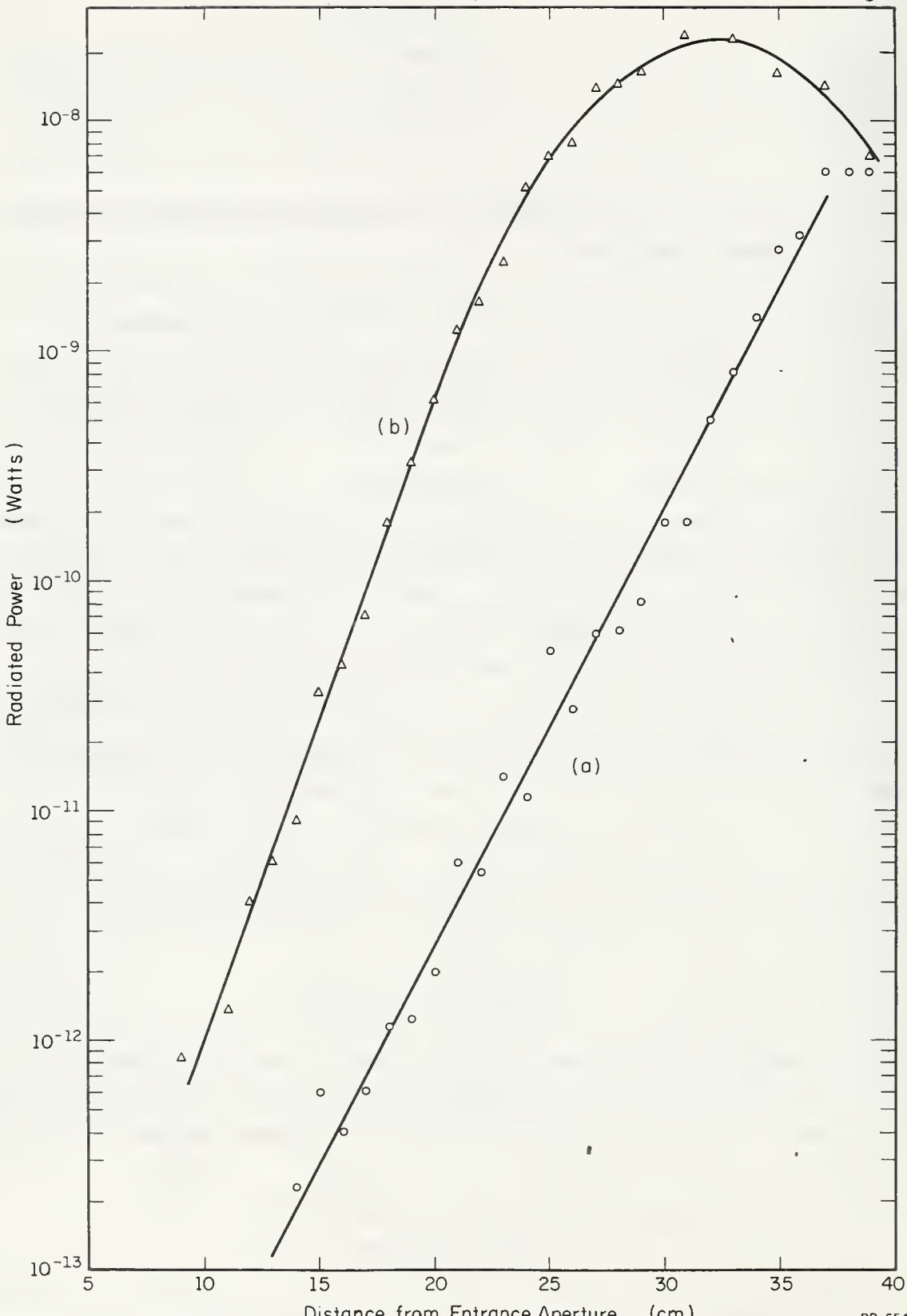


Fig. 3.1. Change of radiated power along the axis of the system. Plasma density $1.2 \times 10^{18} m^{-3}$; (a) beam current: 170 mA, beam density: $1.2 \times 10^{14} m^{-3}$, beam voltage: 18kV; (b) beam current: 290 mA, beam density: $2.2 \times 10^{14} m^{-3}$, beam voltage: 18kV.

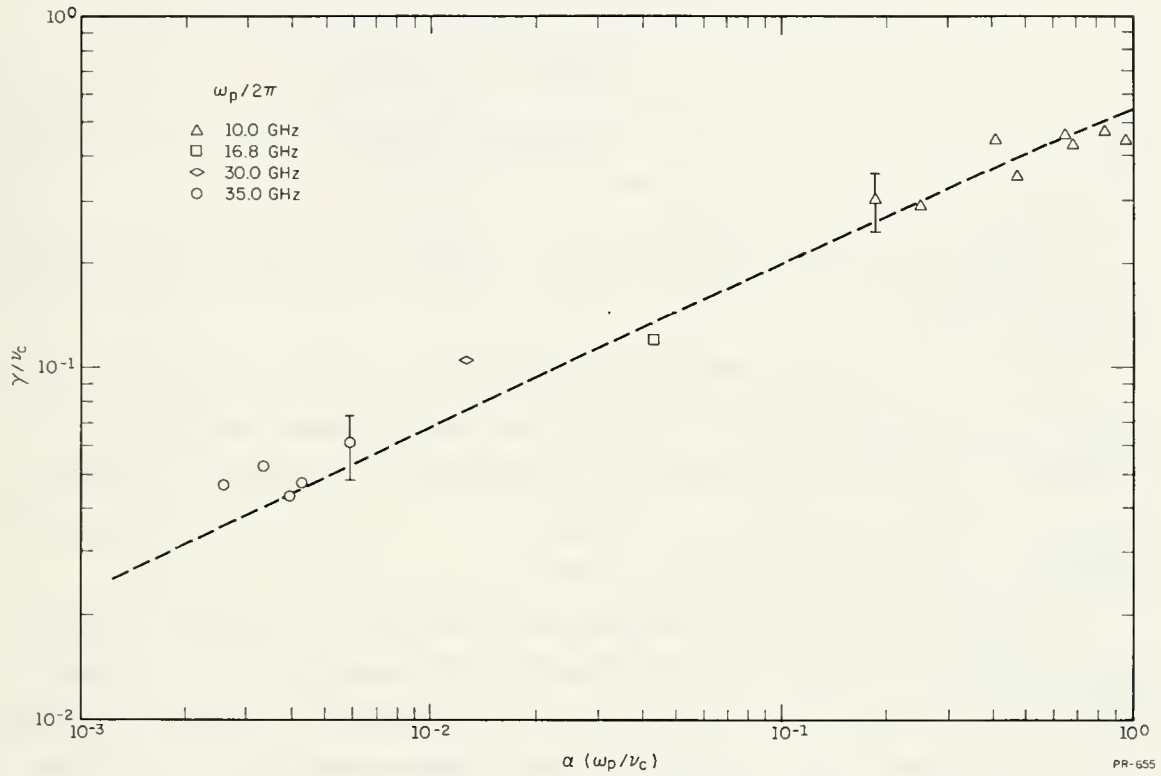


Fig. 3.2. Comparison of theoretical and experimental growth rates for various values of $\alpha(\omega_p/\nu_c)^3$.

to obtain from Eq.(1) to first order

$$(\varepsilon\omega_p/\nu_c)^3 + \tfrac{1}{2}(\varepsilon\omega_p/\nu_c)^2 = \tfrac{1}{2}\alpha(\omega_p/\nu_c)^3 \quad , \tag{2}$$

where $\alpha = \omega_B^2/\omega_p^2 = n_B/n_p$. The growth rate is then given by $\gamma = \mathrm{Im}(\omega) = \omega_p \mathrm{Im}(\varepsilon)$. In the limiting case $\gamma/\nu_c \ll 1$ one finds

$$\gamma/\nu_c = [\tfrac{1}{2}\alpha(\omega_p/\nu_c)^3]^{\frac{1}{2}} \quad , \tag{3}$$

or

$$\gamma = \gamma_o[2^{5/2}\alpha^{1/3}\omega_p/3\nu_c]^{\frac{1}{2}} \quad ,$$

with

$$\gamma_o = (3/4)^{\frac{1}{2}}(\tfrac{1}{2}\alpha)^{1/3}\omega_p \quad , \tag{4}$$

as the collisionless growth rate.

For the electron-ion collision frequency we take the experimental value of Chen [3]

$$\nu_c = 3.6(n_p/T_e^{3/2})\ln(2\times10^7 T_e^{3/2}/n_p^{\frac{1}{2}}) \quad , \tag{5}$$

where T_e is the electron temperature. The plasma temperature in our experiment is 300 K as deduced from the ambipolar diffusion coefficient. Electron-electron collisions can be taken into account by multiplying this expression by 2.4 (Comisar [4]).

The experimental arrangement was described in Ref.[1]. The spatial variation of radiation from the high frequency instability was detected with Ku-1 P- and X-band receivers of 10dB noise figure. Figure 3.1 shows two examples how the radiated power changed along the axis of the beam. As expected, the unstable waves start growing exponentially in space until, at large amplitudes, nonlinear effects become dominant. It is also evident that the waves grow faster for a larger beam current. The final amplitude represents an increase of the fluctuation spectrum by 7 to 8 orders of magnitude above the thermal equilibrium level.

From the initial slopes of curves like the ones shown in Fig. 3.1, we deduced a spatial growth rate σ. One can relate σ to the temporal growth rate γ by $\gamma = \sigma\, v_g$. The group velocity v_g of the most unstable waves can be calculated from the dispersion relation to be $v_g = (2/3)\, v_o$.

Figure 3.2 shows the experimental values of γ/ν_c as a function of the universal parameter $\alpha\,(\omega_p/\nu_c)^3$. The groups of points representing equal plasma frequency were measured for various beam densities. The dashed curve

represents $(\omega_p/\nu_c)\mathrm{Im}(\epsilon)$ as deduced from Eq.(2). We conclude from the comparison of the experimental data with the theoretical curve that Eq.(2) indeed describes the parameter dependence of $\mathrm{Im}(\epsilon)$ correctly in a quantitative way. Especially, the results show for $\gamma/\nu_c \ll 1$ the square root dependence on $\alpha(\omega_p/\nu_c)^3$ as expected from Eq.(3).

In the present experiment, the plasma temperature relaxes to room temperature within about 50 μs after the discharge pulse. This accounts for the high collision frequency at the delay times at which the experiments are performed (100 to 300 μs, corresponding to plasma densities of 2×10^{19} to 10^{18} m^{-3}). We therefore cannot observe the simpler case of zero collision frequency directly. On the other hand we have observed [5] that the beam can heat the plasma and that this heating is a cumulative effect in time and space. In other words, the heating is small at small times and distances and can increase the plasma temperature by up to a factor 10 for long times within the beam and at long distances. According to Eq.(5), an increase in temperature can reduce the collision frequency considerably.

The beam duration is the example of Fig. 3.1 was 6 μs. The data were taken during the first 2 μs of the beam. In Fig. 3.3 this curve a is compared with data taken during the last 2 μs (curve b). Curve b starts with a growth rate which is similar to curve a. At intermediate distances, the growth rate is increased until, at large amplitudes, saturation is reached. The experimental and theoretical growth rates for the conditions of Fig. 3.3 are:

theor. growth rates:	with coll.	1.4×10^9/s
	without coll.	2.6×10^9/s
exper. growth rates:	curve a	1.7×10^9/s
	curve b for $z < 170$mm	1.6×10^9/s
	for $z > 170$mm	2.9×10^9/s

It is evident that the theoretical growth rate including collisions agrees to within the accuracy of the measurements with the values deduced from curve a and with the initial slope of curve b, whereas the collisionless growth rate agrees with the one deduced from the intermediate part of

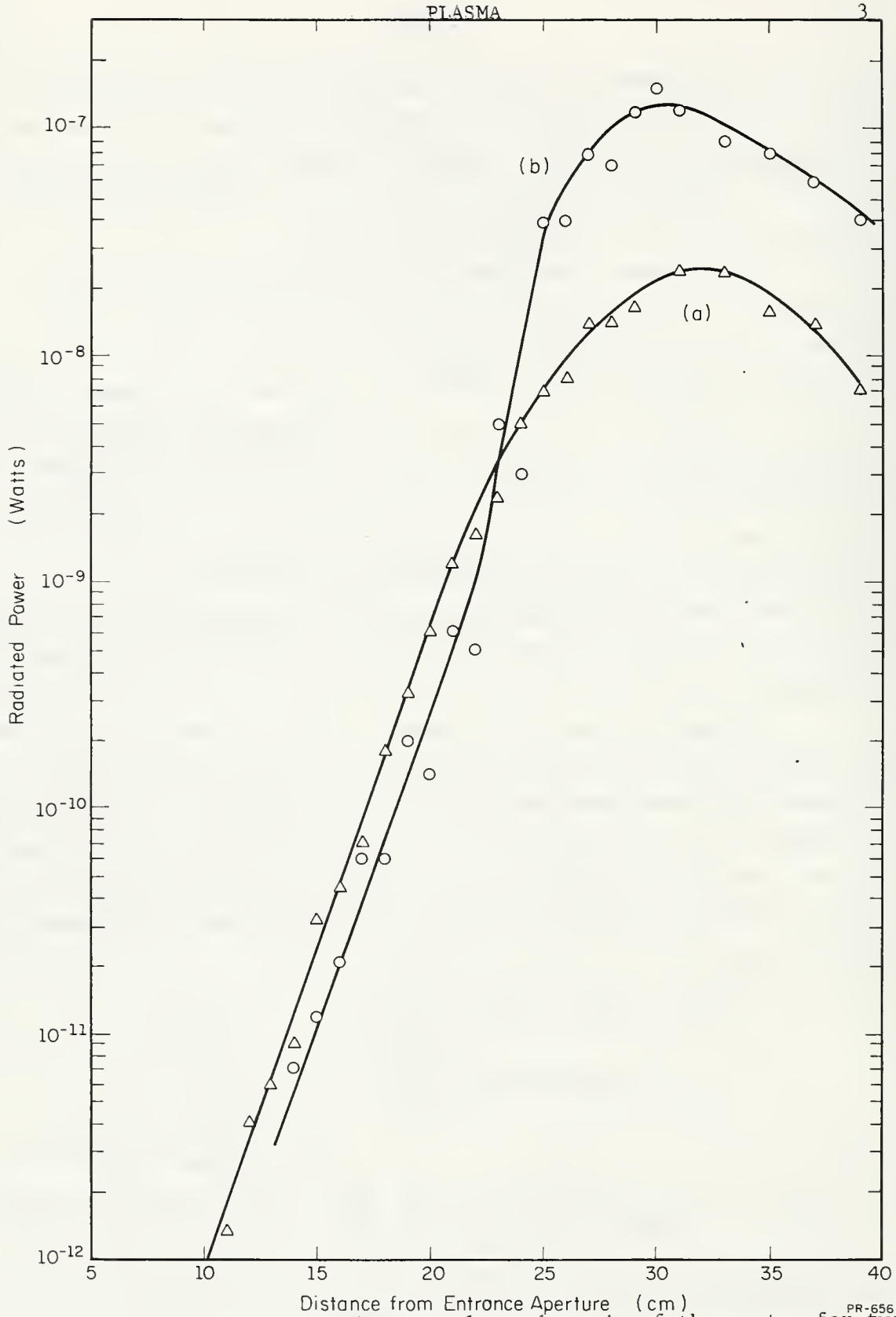


Fig. 3.3. Change of radiated power along the axis of the system for two different times within the beam. Plasma and beam parameters are identical with those of Fig. 1(b); (a) radiated power from 0-2μs, (b) radiated power from 4-6μs.

curve b. We conclude that indeed the collisionless case is reached in this region due to heating of the plasma. It should be noted that the size of this region depends strongly on the beam current. It is not at all observable for tenuous beams.

H. Böhmer
M. Raether

3.2. Influence of Collisions on the Instability of Cold and Warm Electron Beams Interacting with a Plasma

It has been realized for some time that collisions have a different influence on the beam-plasma instability depending on whether the beam is monoenergetic or has a finite velocity spread in excess of a critical value. In the first case, the interaction is nonresonant in the sense that there are no beam particles with velocities equal to the phase velocity of the wave, in the second case the interaction is resonant.

Recent experimental work with beams of finite velocity spread made it desirable to explore in some detail the simultaneous influence of collision and velocity spread on the beam plasma instability, and, in particular, over the transition region between a cold and a warm beam.

To this end,we have studied numerically the dispersion relation of a Lorentzian beam in a plasma with collisions.

The results are compared with previous calculation by Bludman, Watson, and Rosenbluth [2] for the cold beam and those of Ascoli [6] and Singhaus [7] for a Maxwellian beam.

The Lorentzian distribution function for a beam

$$f_o = \frac{\Delta v/\pi}{\Delta v^2+(v-v_o)^2} \quad , \tag{1}$$

in a cold plasma with collisions leads to the dispersion relation

$$1 = \frac{\omega_p^2}{\omega(\omega+i\nu_c)} - \frac{\omega_B^2}{[k\Delta v-i(\omega-kv_o)]^2} \quad . \tag{2}$$

For the collisionless case this dispersion relation has been discussed by O'Neil and Malmberg [8] and Clemmow [9]. It was shown in Ref.[8] that the

results obtained for the Lorentzian beam are in close agreement with those for a Maxwellian beam. The great advantage of the Lorentzian beam is that it leads to a simple algebraic dispersion relation and that, by varying the parameter Δv, the transition from a cold to a warm beam can be explored in detail. Upon introducing the dimensionless variables

$$x = \omega/\omega_p \qquad \beta = \Delta v/v_o$$

$$\varkappa = kv_o/\omega_p \qquad \nu = \nu_c/\omega_p$$

$$\alpha = (\omega_B/\omega_p)^2$$

Eq.(2) goes over into

$$1 = \frac{1}{x(x+i\nu)} - \frac{\alpha}{[\varkappa\beta - i(x-\varkappa)]^2} \quad . \tag{3}$$

Cold beam:

For $\beta = 0$, Eq.(3) reduces to that for a cold beam, which was treated in Ref.[1]

$$1 = \frac{1}{x(x+i\nu)} + \frac{\alpha}{(x-\varkappa)^2} \quad . \tag{4}$$

In order to clarify the mode structure and the nature of the collisional enhancement at low frequencies, we have solved this equation for selected values of the parameters. In all the following examples α was kept constant at 10^{-4}. This value was chosen because it is representative of a typical experimental situation.

Figure 3.4 shows a plot of the growth rate γ/ω_p as a function of $\varkappa$ for various values of ν. We notice that with increasing collision frequency, the maximum unstable wavenumber, which for the collisionless case is given by

$$\varkappa_{opt} = 1+(3/2)(\alpha/2)^{2/3} \quad , \tag{5}$$

shifts to lower values. At the same time, the unstable wavenumber range, which for the collisionless beam extends from $\varkappa = 0$ to $\varkappa = (1 + \alpha^{1/3})^{3/2}$, extends out to $\varkappa \to \infty$. The maximum growth rate goes asymptotically towards a constant value in the limit of large collision frequency. This asymptotic value is readily obtained from the dispersion relation in the limit $\nu >> x$. Eq. 4 then reduces to

$$1 = 1/i\nu x + 1/(x-\varkappa)^2 \quad .$$

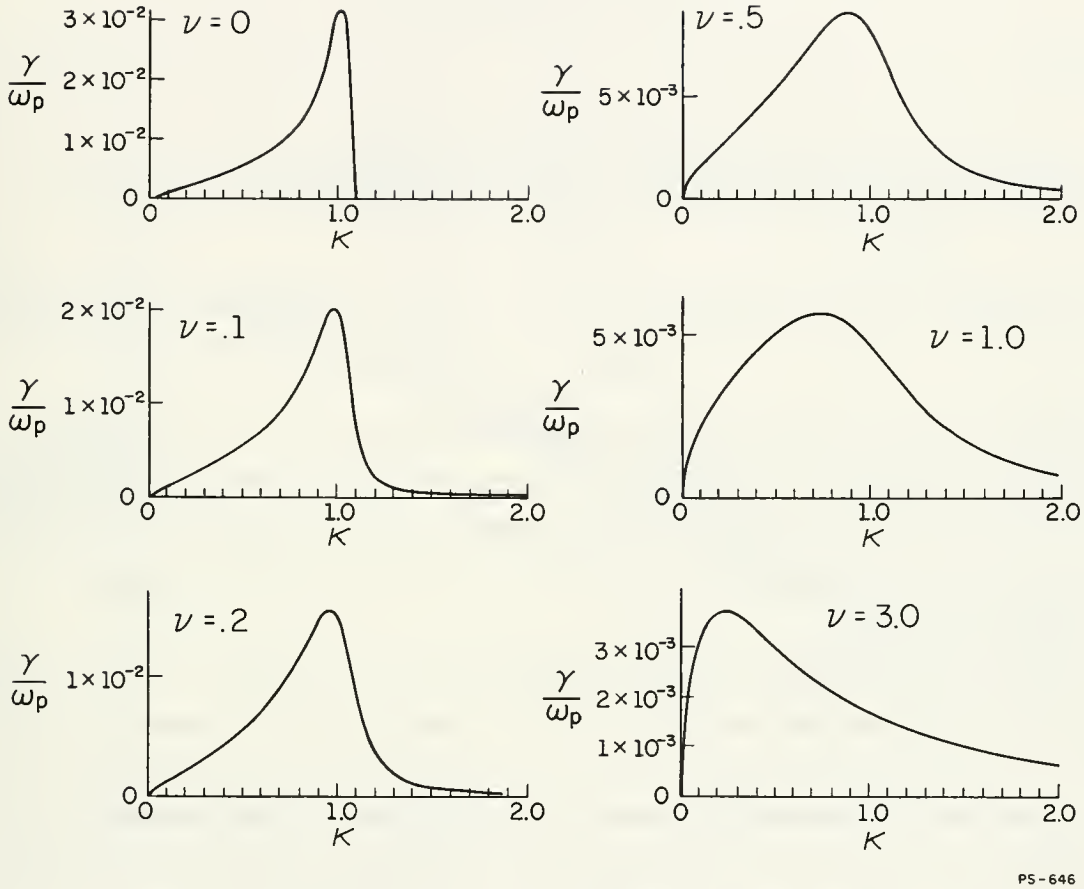


Fig. 3.4. Growth rate as a function of wavenumber for different collision frequencies ($\alpha = 10^{-4}$).

On the unstable branch we put $x = \varkappa + \epsilon$ where $\epsilon \ll 1$ and find to lowest order in $\epsilon/\varkappa$

$$1 = i/\nu\varkappa + \alpha/\epsilon^2 .$$

For the imaginary part of ϵ we find

$$\mathrm{Im}\epsilon = (\alpha/2)^{\frac{1}{2}}[\xi^{\frac{1}{2}} - \xi]^{\frac{1}{2}} , \tag{6}$$

where

$$\xi = (\nu\varkappa)^2/(1+\nu^2\varkappa^2) . \tag{6a}$$

The maximum growth rate follows from the condition

$$d\mathrm{Im}\epsilon/d\varkappa = 0 .$$

This condition is satisfied for $\varkappa_{opt} = 1/(\nu/3)$; the growth rate follows as

$$(\mathrm{Im}\epsilon)_{opt} = \tfrac{1}{2}(\alpha/2)^{\frac{1}{2}} . \tag{7}$$

In Fig. 3.5, we have plotted the maximum growth rate as a function of collision frequency together with the approximate expression from Ref.[2]. The large collision frequency approximation of BWR agrees to within 10% with the exact solution in the experimentally interesting region $0.1 < \nu < 0.5$.

The breakdown of this approximation for larger ν is due to the fact that the maximum unstable $\varkappa$ shifts away from the assumed value $\varkappa = 1$.

Figure 3.6 shows a plot of the real and imaginary parts of the roots of Eq.(2) in the neighborhood of $\varkappa = 1$, with ν as a parameter. The picture is reminiscent of the behavior of a warm beam without collision as discussed by O'Neil and Malmberg [8]. The essential difference is that the unstable root is always located on the "slow-beam branch".

Although the present example refers to only one particular value of α, it is easy to show that, in the neighborhood of $\varkappa = 1$, $\gamma/\omega_p\nu$ is a universal function of α/ν^3.

The physical mechanism of the low-frequency instability can be understood easily if we observe that for $\nu_c \gg \omega$, the equation of motion for the plasma electron gives a pure viscous drift velocity of

$$v_D = (e/m)(E/\nu_c) .$$

Consider a charge perturbation of wavelength λ on the beam. The electric field associated with a charge density n_B' is $E \approx 4\pi e n_B' \lambda$. This field gives rise to a drift velocity of the plasma electron of

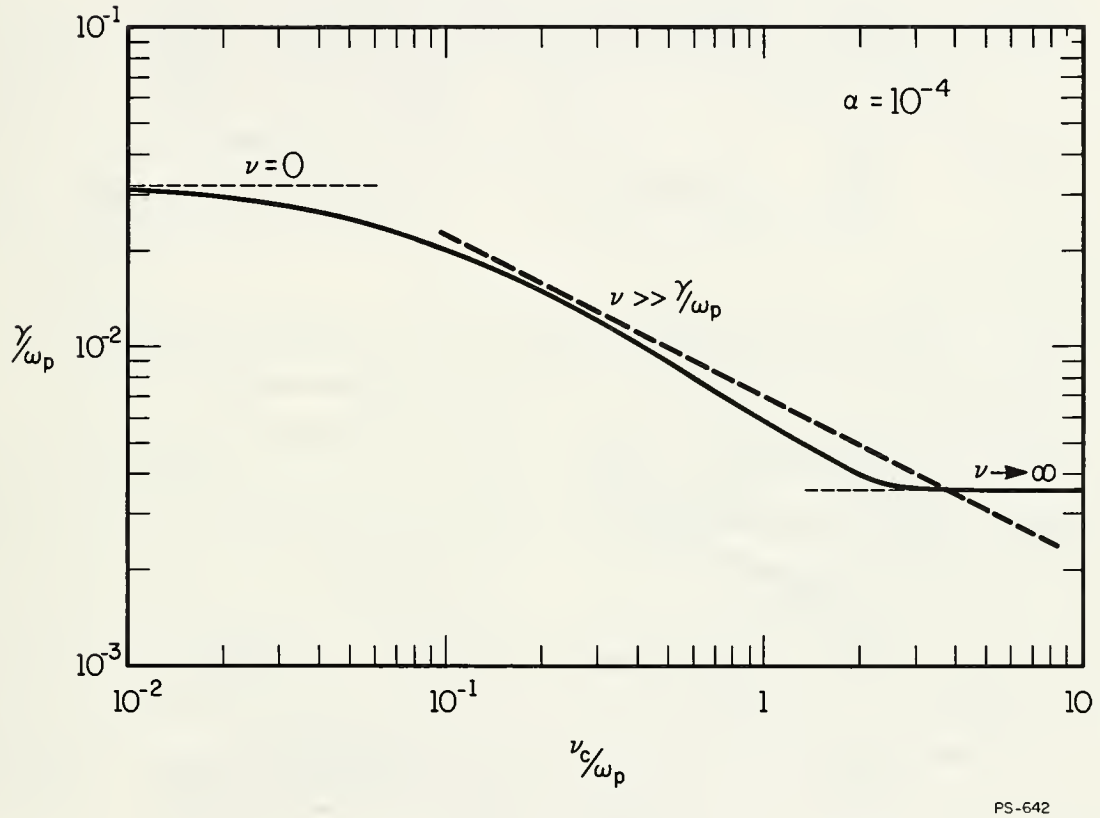
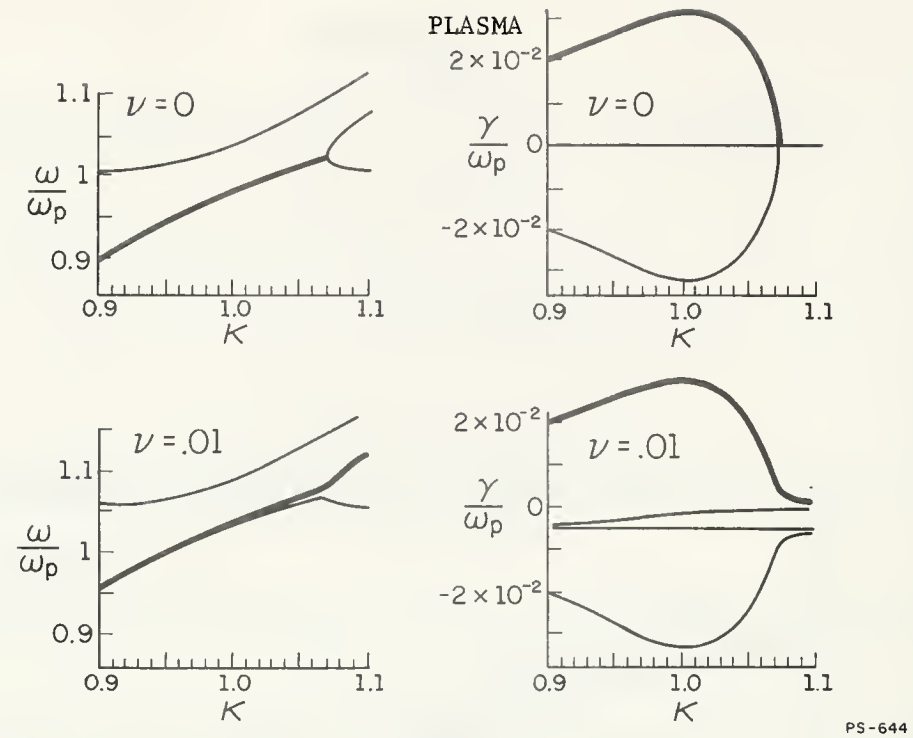


Fig. 3.5. Maximum growth rate for the cold beam instability as a function of collision frequency ($\alpha = 10^{-4}$).



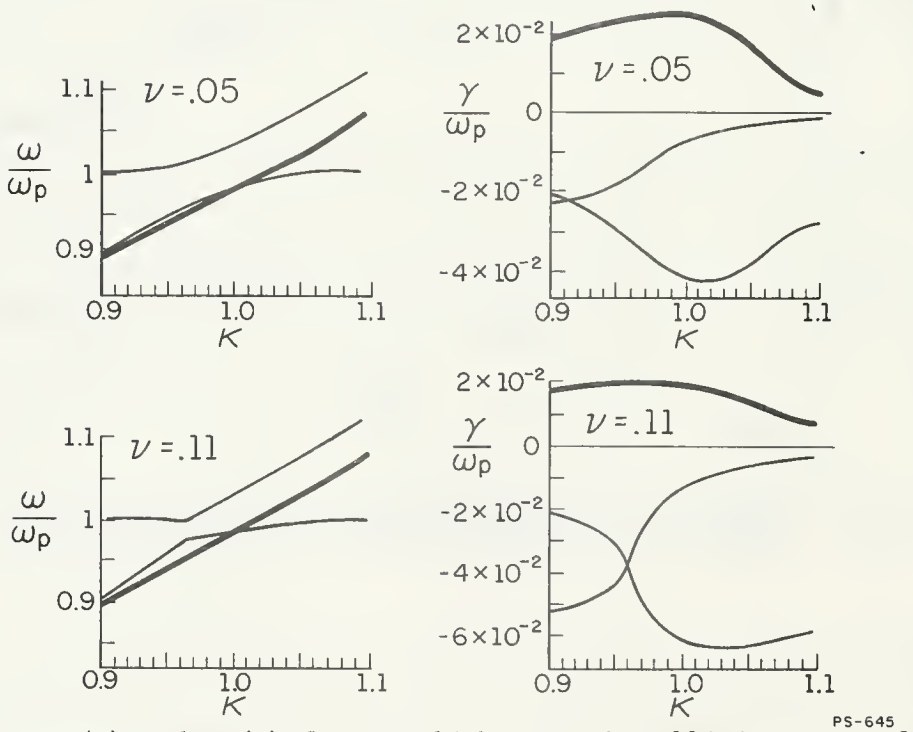


Fig. 3.6a. Re(ω) and Im(ω) for a cold beam with collisions as a function of wavenumber ($\alpha = 10^{-4}$).

$$v_D = (e/m)(4\pi e n_B' \lambda/v_c) .$$

The plasma electrons move in such a fashion as to compensate the beam space charge. The time necessary to do this is

$$\tau = (n_B'/n_p)(\lambda/v_D) = \nu_c/\omega_p^2 .$$

The fraction n_B'/n_p appears because only this fraction of the plasma electron is necessary for space charge neutralization. If this time is equal to $1/\omega$, the neutralizing space charge is just out of phase with the perturbation and the original perturbation is enhanced. This gives the instability condition $\nu_c/\omega_p^2 = 1/\omega = 1/k v_o$ or $\nu\varkappa = 1$, which is of the correct order of magnitude. The low-frequency instability can therefore be classified as a viscous-drag instability, similar in nature to the hose instability.

3.3. Simultaneous Action of Collisions and Velocity Spread

The effect of a velocity spread in the absence of collisions on the instability of a Lorentzian beam has been treated extensively by O'Neil and Malmberg [8]. The somewhat surprising result of their work is that, for a critical velocity spread, the unstable mode shifts from the "beam branch" to the "plasma branch" of the dispersion relation. If collisions are present in addition to a velocity spread, the question arises as to how the transition from the unstable beam branch to the unstable plasma branch proceeds when the system goes from a collision-dominated to a velocity-spread-dominated situation. Calculations for $\alpha = 10^{-4}$, $\nu = 0.01$ show that, as the velocity spread is turned on, the unstable region on the beam branch shrinks from $\varkappa = \infty$ to a finite value $\varkappa_{max}$. With increasing velocity spread $\varkappa_{max}$ decreases until for $\beta = \nu/2$ the roots of the dispersion relation take on a shape reminiscent of the cold-beam dispersion relation except that all roots are offset by $-i\nu/2$. If the velocity spread is increased further, the unstable mode shifts over to the plasma branch, and the subsequent behavior is qualitatively similar to the collisionless case.

The velocity spread has moreover the effect of splitting the instability into two distinct modes, one at high and one at low frequencies. This is particularly noticeable at high collision frequencies.

Figure 3.7 shows an example for this behavior. These two modes behave differently as a function of velocity spread. Whereas the low-frequency mode has always finite, though small, growth rates, the high-frequency mode damps out above a certain critical velocity spread. This quenching of the high frequency mode has been noticed before. Ascoli [6] and Singhaus [7] derived a criterium for the quenching of the instability of a Maxwellian beam under the condition $\Delta v/v_o >> \alpha^{1/3}$ (warm beam).

In our notation, the Singhaus criterium states that the high-frequency mode is quenched provided that

$$\nu\beta^2/\alpha > 0.66, \tag{8}$$

and

$$\beta >> \alpha^{1/3}.$$

In Fig. 3.8, we have plotted the maximum growth rate as a function of ν and β for $\alpha = 10^{-4}$. We see that the Singhaus criterium as given by Eq.(8) agrees very well with the zero growth-rate contour for the Lorentzian beam and that it is actually valid for much lower velocity spread than the condition $\beta >> \alpha^{1/3}$ would indicate.

For collision frequencies $\nu > 0.6$ the distinction between low- and high-frequency mode loses its meaning, and there is a continuous shift from high to low frequency as β is increased.

This behavior is further illustrated in Fig. 3.9 where the unstable wavenumber range is plotted as a function of β for different values of ν as a parameter.

We note that the low-frequency mode is always present for finite collision frequency, although it is restricted to a very small wavenumber range.

M. Raether

3.4. Experiment with Beams of Finite Velocity Spread

It was shown in the previous section that collisions have a drastically different effect, depending on whether the beam is cold or warm. We have investigated the instability of electron beams of finite velocity spread, interacting with a neon afterglow plasma in the density regime where coulomb collisions are important. The velocity spread of

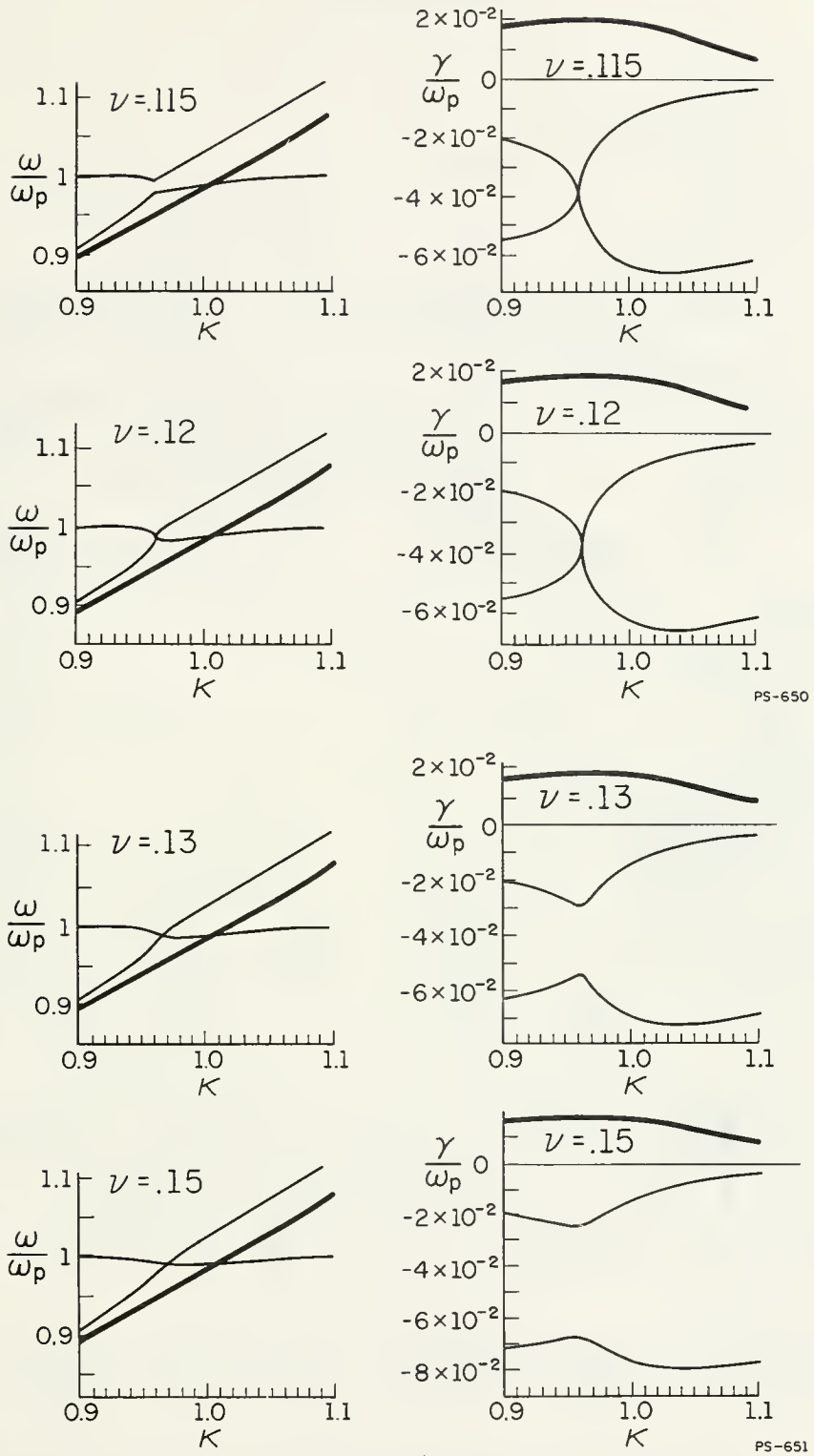
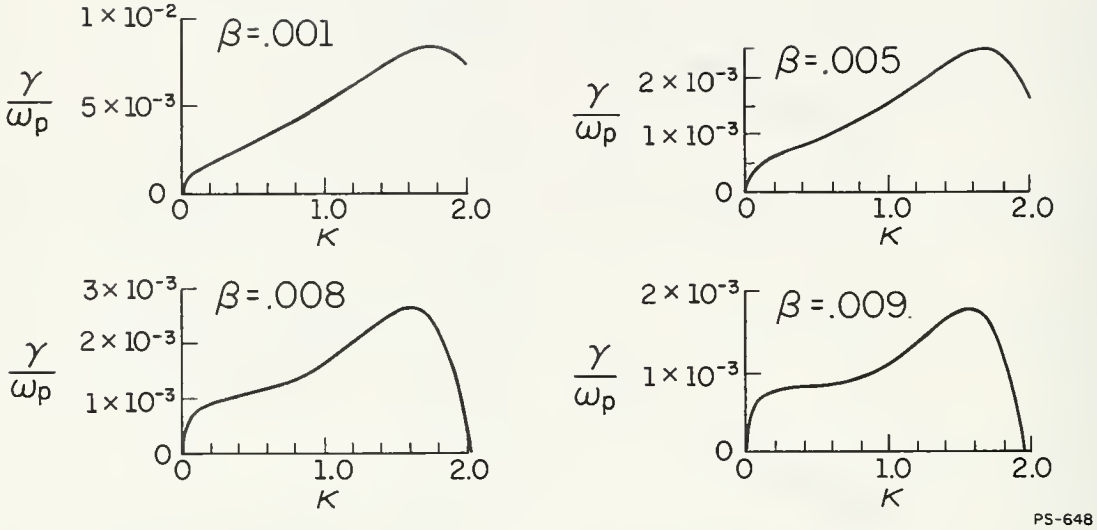


Fig. 3.6b. Continuation of Fig. 3.6a.



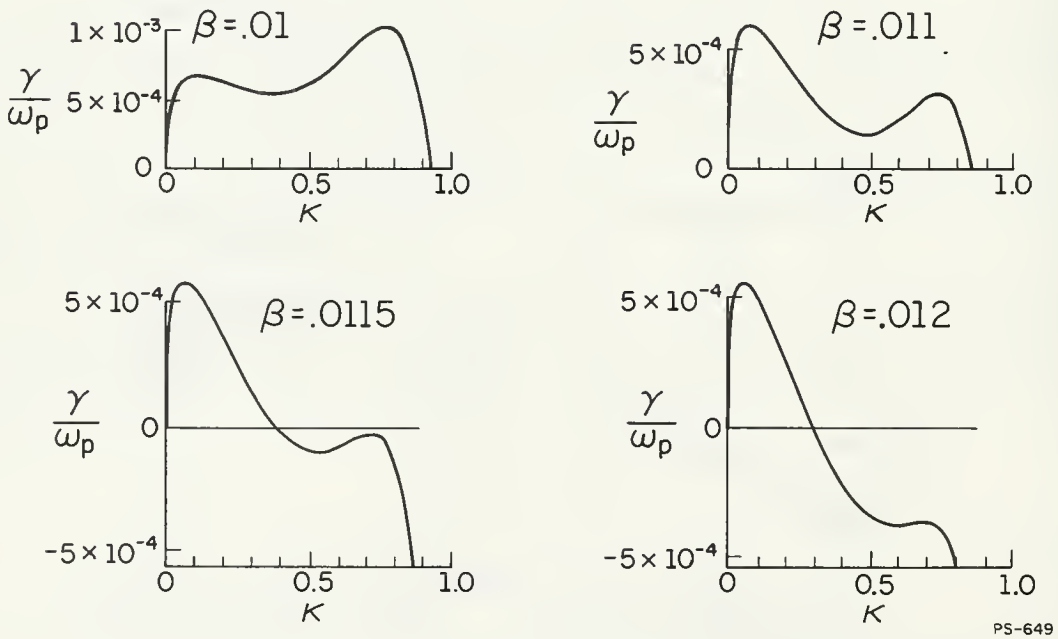


Fig. 3.7. Splitting into high and low frequency mode ($\alpha = 10^{-4}$).

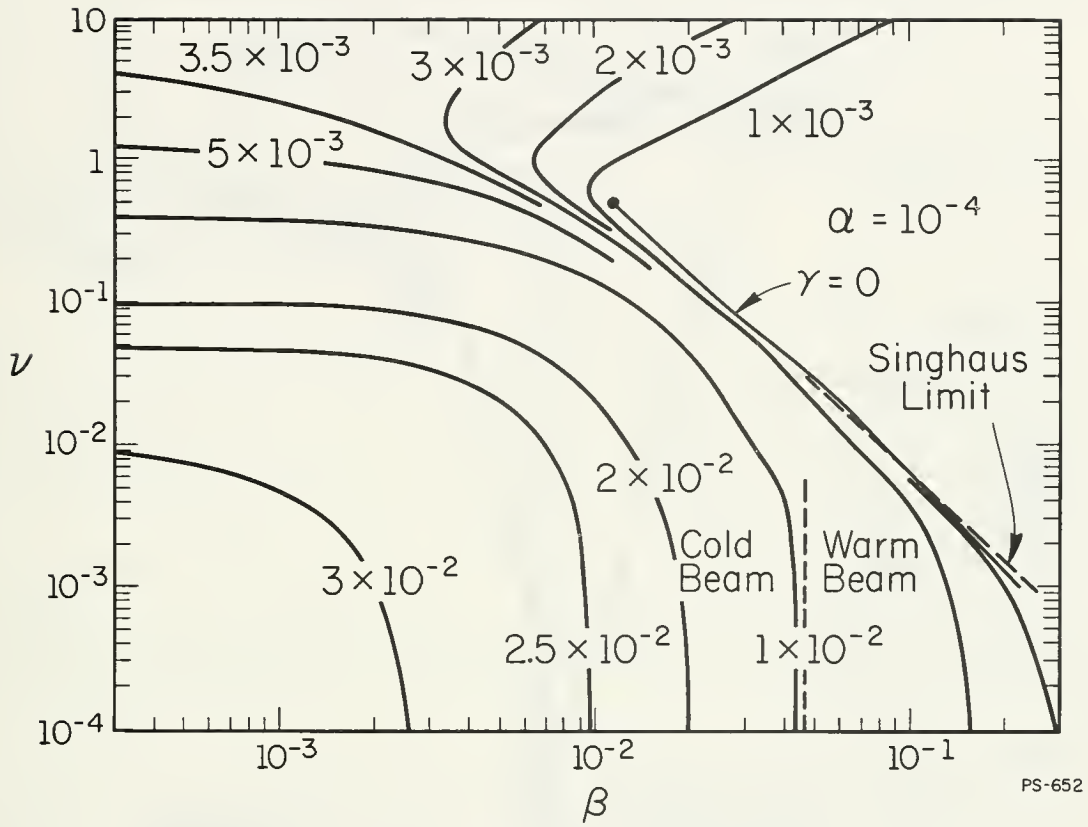


Fig. 3.8. Contours of equal growth rate in the ν-β plane. The indicated parameters are γ/ω_p ($\alpha = 10^{-4}$).

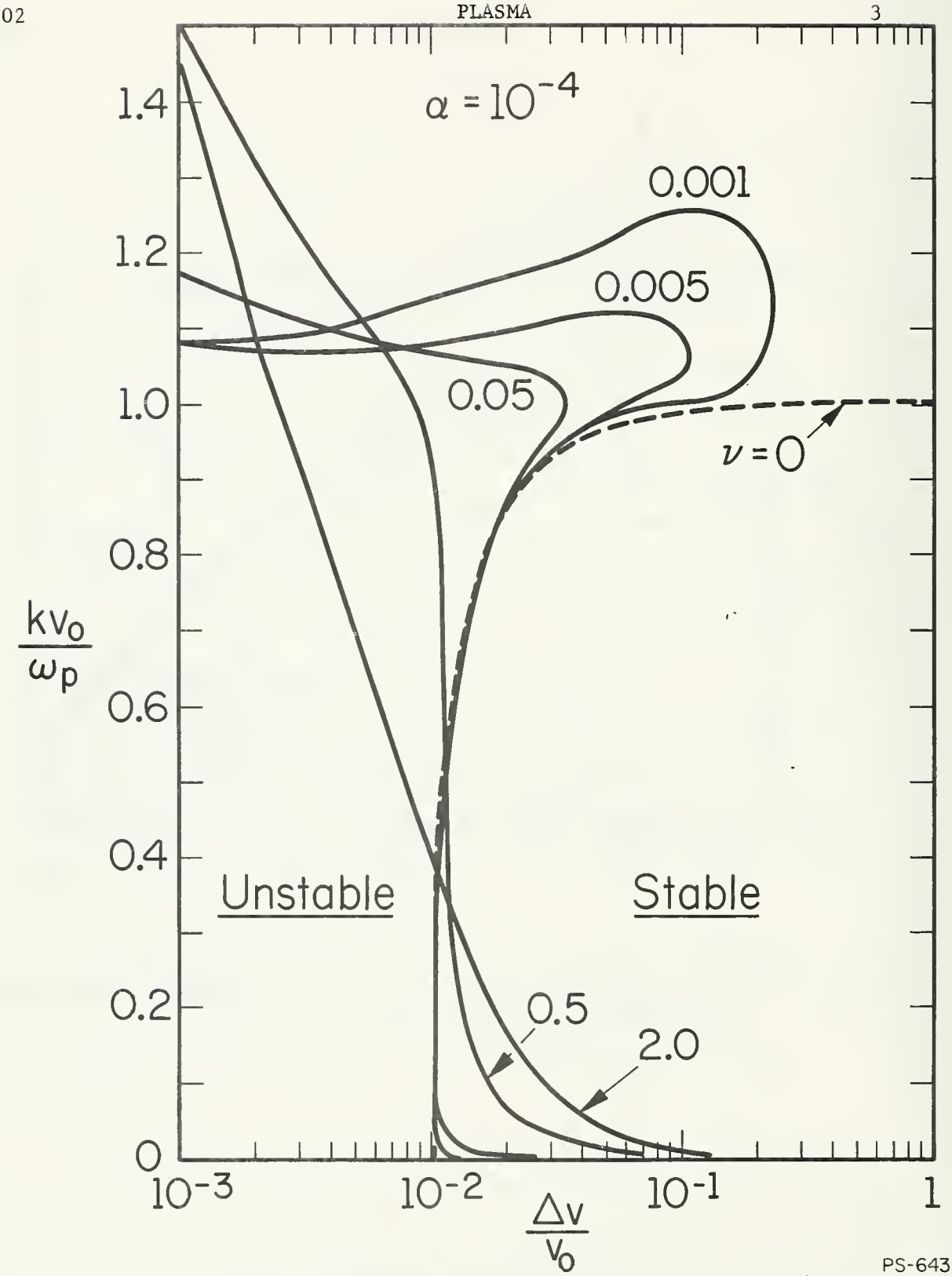


Fig. 3.9. Boundary between stable and unstable region in the β-$\varkappa$ plane ($\alpha = 10^{-4}$).

the beam was produced by passing the beam through Beryllium foils of approximately 1μm thickness [10]. Values of u/v_o of 1-2% were produced with transmitted beam currents up to 400 mA at energies between 14 and 17 kV, 1 to 2 μs beam duration and 15 Hz repetition frequency. The energy distribution was measured with an electrostatic deflection type analyzer of 0.3% resolution. Measurements were performed at plasma densities of $1.2\times10^{18}/m^3$ and $1.4\times10^{19}/m^3$. The instability was detected by measuring the microwave radiation from the unstable plasma waves with receivers of 10 dB noise figure in the X and K band and by monitoring the beam distribution function at the end of the interaction-region.

For cold beams the instability shows a spatial growth with a growth rate $\sigma = 2\gamma_c/v_o$. The radiated power increases by 6 to 9 orders of magnitude over the interaction region, depending on beam density. As the velocity spread becomes larger, the growth rate, the radiated power and the velocity spread produced by the interaction decrease. For values of $\nu\beta^2/\alpha^{1/3} > 0.66$ (Singhaus limit) radiation is no longer detectable, that is, the radiated power is less than 10^{-13} W, which is the sensitivity limit of our receivers. At the same time, the change in the distribution function produced by the interaction decreases rapidly and becomes undetectable for values of $\nu\beta^2/\alpha^{1/3} \gg 1$.

In Fig. 3.10 the ratio $(u)_{final}/(u)_{initial}$, which is a convenient relative measure for the growth rate and in many cases more sensitive than the radiated power, is plotted for a variety of cases versus $\nu\beta^2/\alpha^{1/3}$. We see that this ratio goes to 1 for values close to the Singhaus limit.

The quantity $\beta/\alpha^{1/3}$ is still not large compared to unity but ranges between 0.3 and 1.0. Nevertheless, the effect of collisions on the instability is already quite dramatic. As a demonstration for this, we present one particular example in Table 3.1 where a cold and a warm beam are compared. The beam and plasma densities and the collision frequency were the same in both cases, only the velocity spread was different. The plasma density was $1.4\times10^{19}/m^3$, radiation was detected at 35 GHz. The symbols are I_B for the beam current, V for the beam energy, and P for the maximum radiated power. As a consequence of the velocity spread the microwave power radiated by the instability has decreased by more than 3½ orders of magnitude.

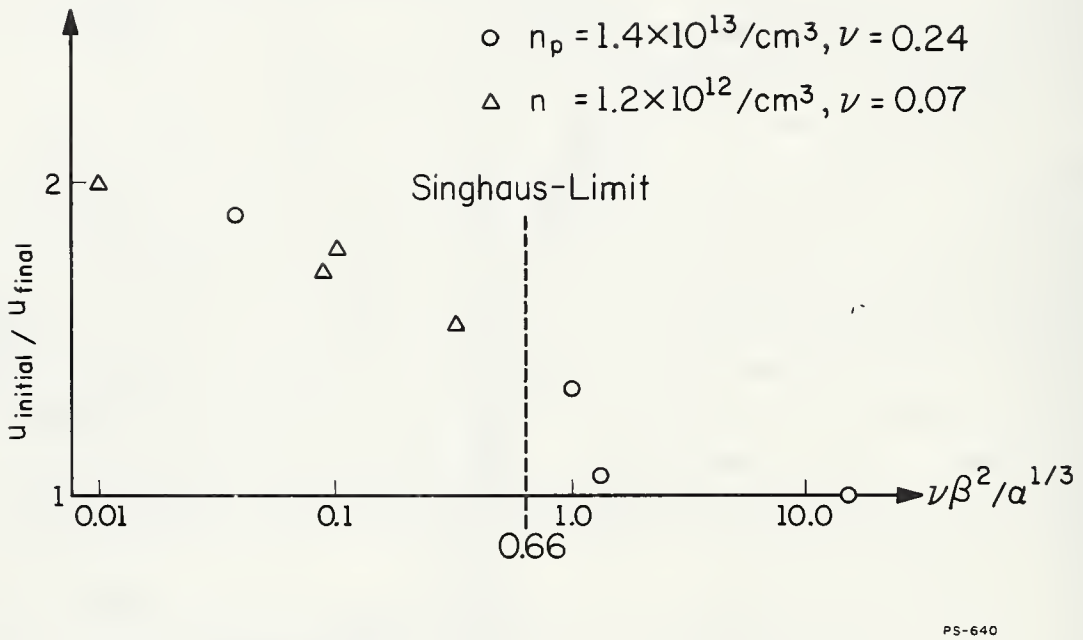


Fig. 3.10. Decrease of instability as a function of velocity spread.

Table 3.1

I_B	V	n_B	$(u/v_o)_i$	$(u/v_o)_f$	P	$\nu\beta^2/\alpha^{1/3}$
320 mA	17 kV	$3\times10^{14}/m^3$	<0.2%	0.4%	2×10^{-10}W	0.042
400 mA	16 kV	$3.1\times10^{14}/m^3$	1%	1.3%	$< 10^{-13}$W	1.05

We conclude that collisions have a much larger effect on the instability of beams with finite velocity spread than on monoenergetic beams; even under conditions where $\beta \leq \alpha^{1/3}$. For $\beta \approx \alpha^{1/3}$ and $\nu\beta^2/\alpha > 0.66$, the instability is effectively suppressed. Although the distribution functions used in these experiments were not displaced Maxwellians, the experimental results compare favorably with the theoretical model calculated by Singhaus and the calculations presented in the previous section.

J. Chang
H. Böhmer
M. Raether

3.5. Dispersion Relations and Mixed Boundary Value Problems

Systems of linear partial differential equations, used to study the stability of plasmas, are frequently examined by considering plane wave solutions $\exp[i(kx-\omega\tau)]$, and the resulting dispersion relation, $D(k,\omega) = 0$, required for a non-trivial solution. Previous studies of the physical significance of the roots of $D(k,\omega) = 0$ have commonly been limited to either initial value problems (infinite in space) or boundary value problems (infinite in duration). The present investigation is concerned with the influence of mixed boundary conditions, appropriate for semi-infinite temporal-spatial regions, on the character of the instabilities. It has become important to clarify this situation in order to interpret beam-plasma instabilities, which involve both boundaries (the injection point of the beam) and the initial perturbations within the plasma.

The analysis of the instability produced by a monoenergetic beam injected into a cold plasma has been generalized to include the collision of the plasma electrons with background ions. The complicated space-time development of the instability, becomes relatively simple only in the asymptotic limit $t \gg x/u_o$ (x is the distance from the injection point;

u_o is the beam velocity). In this limit, the electric field, arising from a single perturbation at t = 0, behaves essentially as $\exp\{(3\sqrt{3}/4)(x\omega_v/u_o)^{2/3} \times(\omega_p t)^{1/3}-\frac{1}{2}\nu\tau\}$ where (ω_b,ω_p) are the (beam, plasma) plasma frequencies, and ν is the collision frequency (and $\omega_p \gg \nu$ has been assumed). This function has a maximum for fixed x when $t \doteq (\frac{1}{2}\sqrt{3})^{3/2}(\omega_p/\nu^3)^{\frac{1}{2}}(\omega_b x/u_o)$, which is in the asymptotic region ($t \gg x/u_o$) only if $\omega_p\omega_b^2/\nu^3 > 1$. If this inequality is not satisfied the above expression reduces to $\exp(-\nu t/2)$, which means that the instability is simply damping in the asymptotic region. In many experiments, the condition $\omega_p\omega_b^2/\nu^3 > 1$ is not satisfied, in which case it is necessary to examine the instability in the region $t \approx x/u_o$ in order to obtain interesting results (e.g., the maximum amplitude of the instability). A study of the solution in this region is now being carried out.

To clarify a number of more general aspects of mixed boundary value problems, an analysis has been made of the generalized wave equation

$$[(\partial/\partial t+c_1\partial/\partial x)(\partial/\partial t-c_2\partial/\partial x)-b]u(x,t) = 0.$$

This equation can have any of four common forms of dispersion relations (depending on the signs of b, c_1, c_2), yet is of low enough order to make analysis relatively simple. A detailed study of each of the four cases has been made, which illustrate (i) limitations on the conclusions one can infer from the dispersion relation, (ii) the inter-relationship between characteristics, plane waves, and boundary conditions, (iii) the role of retarded boundary conditions in physical (vs mathematical) discussions of instability, and (iv) problems concerning the asymptotic stability ($t \to \infty$) of finite systems, for which no usable mathematical method of analysis appears to be available.

E. A. Jackson

3.6. References

1. CSL Progress Report, March-August 1967.
2. S. A. Bludman, J. M. Watson and M. N. Rosenbluth, Phys. Fluids 3, 747, 1960.
3. C. L. Chen, Phys. Rev. 135, A627 (1964).
4. G. G. Comisar, Phys. Fluids 6, 76 (1963).
5. CSL Progress Report, September 1967 through February 1968.

6. G. Ascoli, CSL Report R-131, Urbana, 1961.
7. H. E. Linghaus, Phys. Fluids 11, 1754 (1968).
8. T. M. O'Neil and J. M. Malmberg, Phys. Fluids 11, 1754 (1968).
9. P. C. Clemmon, J. Plasma Physics 2, 85 (1968).
10. J. Chang, H. Böhmer, M. Raether, Rev. Sci. Instr. 39, 1873 (1968).

4. RAREFIED GAS DYNAMICS†

S. M. Yen W. Chow A. Nordsieck*
B. L. Hicks W. H. Chu B. J. Reilly

4.1. The Boltzmann Computer Program

A plan has been initiated to incorporate in our computer program more realistic gas-surface interaction models. Since our Monte Carlo results yield distinct nonequilibrium behavior in the Knudsen layer for the heat-transfer problems, it would be of interest to study further this behavior by using more realistic boundary conditions. Just like the boundary layer in continuum fluid dynamics, the Knudsen layer encloses an important flow region. As a first step of our plan to use appropriate boundary conditions, we have incorporated in our computer program a model with specular as well as diffusive reflections to study the difference in the nonequilibrium behavior in the Knudsen layer for heat-transfer problems. A FORTRAN program permits the generation of appropriate distribution functions to be used as those of the zeroth iterate for the iterative integration scheme.

We have also started our plan to apply Nordsieck's Monte-Carlo method to more complex problems by using ILLIAC IV, a large-scale, parallel-operation computer to be installed at the University of Illinois in 1970. The initial phase of this plan is to rewrite our Monte-Carlo program for ILLIAC IV. We have rewritten two Monte-Carlo programs for evaluating the Boltzmann collision integrals for ILLIAC IV, one in ALGOL language and the other in TRANQUIL (an assembly language for ILLIAC IV). The ALGOL program is now being tested on the Burroughs B-5500 computer. The latter computer is now used to simulate the parallel operation of ILLIAC IV. The test now being conducted is to input a fixed set of random numbers and to examine the collision integrals. Sixty-four processing elements of ILLIAC IV are scheduled to be installed in the fall of 1970.

A program is now being written to compute analytically the collision integral for the Mott-Smith distribution function in a shock wave.

---

†This work was supported by the Office of Naval Research under Contract No. ONR N00014-67-A-0305-0001; auxiliary support was provided by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199.

*Consultant

The results on the collision integrals will be used to make detailed comparison with our Monte-Carlo values of the collision integrals and to ascertain further the accuracy of our calculations. We have already found that the isolines of the collision integrals for $M_1=3$ are in good agreement with the analytical results obtained by Narasimha.

4.2. The Boltzmann Shock Wave

The solutions of the Boltzmann equation we have obtained for a large range of Mach numbers (1.05 to 10) yield the distribution function, the collision integral, and their functionals. We have analyzed many of these functionals as well as the microscopic properties that reflect the detailed nonequilibrium behavior in a shock wave. The results of this analysis will be summarized for journal publication.

Recently Muntz of General Electric has successfully measured certain distribution functions ($\int f du dv$ and $\int f dv dw$, for f the distribution function, and u the velocity in the direction of flow) in a shock wave of M=1.56 and found his results to be in agreement with the Navier-Stokes results. We have obtained the Monte Carlo solutions for the shock wave for the same Mach number (M=1.56) and made detailed comparison with the corresponding results of the Navier-Stokes shock. For one of the distribution functions measured by Muntz, namely $\int f dv dw$, the Monte Carlo result was also found to be in good agreement with the Navier-Stokes results.

We have made a direct test of the uniqueness of our solution of the Boltzmann equation by using the following initial distribution in our iterative scheme: the upstream equilibrium distribution function f_1 for the "cold half" ($\hat{n} = (n-n_1)/(n_2-n_1)$ less than $\frac{1}{2}$) of the shock, the downstream equilibrium distribution function f_2 for the "hot half" ($\hat{n}$ greater than $\frac{1}{2}$), and $(f_1+f_2)/2$ for the center of the shock ($\hat{n} = \frac{1}{2}$). This distribution function is as close to a step function as we can use in our program. Tests have been applied to the Krook shock, which uses the same integration scheme as the Boltzmann shock, and the Boltzmann shock for M=2.5 and 4.

Some of the results on the uniqueness test for the Krook shock, are summarized as follows: (a) M = 2.5: To reach a value of 0.01% for the percent residual δf (rms value of the difference of the distribution functions

of successive iterations divided by the average value of the distribution function), the number of iterations increases from 15 to 30 when the initial distribution functions f_o are changed from the Mott-Smith f_{MS} to the step function. For the two corresponding solutions at this level of δf, the difference is less than 1% for the density gradient dn/dx and less than 0.1% for $\mathcal{M}_9$ ($\mathcal{M}_9 = \int v_\perp^2 f d\underset{\sim}{v}$) which is used to calculate the temperature distribution. (b) M = 4. To reach a value of 0.1% for percent δf, the number of iterations increases from 12 to 22 when f_o is changed from f_{MS} to the step function. For the two corresponding solutions, the difference is less than 0.2% for dn/dx and less than 0.002% for $\mathcal{M}_9$. For the Boltzmann shock, to reach the value of 0.02% for percent δf, the number of iterations increases from 5 to 36 for M = 2.5 and from 12 to 28 for M = 4 when f_o is changed from f_{MS} to the step function. For the two corresponding solutions for M = 2.5 and 4, the difference in dn/dx and that in $\mathcal{M}_9$ are less than the 90% confidence limits.

The error in df/dn due to the algorithm used in the integration scheme has been studied for the shock for M = 2.5, 4, and 10. The fractional error in df/dn was found to be generally less than 0.01 for the three values of M throughout the velocity space except where one might expect large errors, e.g., near extrema of the distribution function f.

The variable of integration used in our iterative scheme for the shock wave is the number density n. The integration interval between successive stations is $\Delta n_{anal} = (n_2-n_1)_{anal}/(J-1)$, in which $(n_2-n_1)_{anal}$ is the range of integration calculated analytically and (J-1) is the number of stations. Because of quadrature errors, the range of integration (n_2-n_1) calculated numerically in the program is not equal to the analytical value; therefore, certain errors exist due to the use of the analytical value. We conducted the study of such errors by comparing the solutions using Δn_{quad} with those using Δn_{anal} for M = 1.2, 2.5, 6, and 10. We have found that the percent differences between the moments of the distribution function and insignificant.

4.3. The Pseudoshock

More accurate results have been obtained for the initial behavior of the pseudoshock for M = 4 and 6 by using the new techniques of 1S (one sample) and the MB (Maxwell-Boltzmann) correction. Comparisons have been made with the results obtained previously.

4.4. Heat-Transfer Problems

Earlier this year, we have found that the MB correction, one of the improvements of our method developed to extend the application of our Monte-Carlo method for gases very near thermal equilibrium, was necessary in studying the nonequilibrium behavior in the Knudsen layer which was found to be more distinct for smaller Knudsen numbers. The effect of this correction on the heat-transfer solutions for larger values of Knudsen number has been studied.

Because of the MB correction developed, we are able to obtain results for much lower Kn, i.e., in the near-equilibrium flow regime. We have obtained the following two sets of runs that have the same parameters as those used by Springer to obtain the density distribution experimentally; (1) Kn=0.248, T_1/T_2=0.7826, α=0.826, and (2) Kn=0.0549, T_1/T_2=0.7826, α=0.826. (Previously we have made a comparison with his experimental results for Kn=0.6313, T_1/T_2=0.7826, and α=0.826.) Again, it was found that the density distributions are in good agreement with the experimental results except in the region near the hot plate for the case with Kn=0.0549, and that several forms of non-equilibrium behaviors in the Knudsen layer for the two cases are in distinct contrast to those outside this layer.

In the MIT paper, we have presented a part of the results on macroscopic properties for the heat transfer problems. The detailed results on other macroscopic properties, as well as the microscopic properties, are being analyzed. More accurate results have been obtained by using the new technique of 1S as well as the MB correction. Some of these results will be presented in the ASME Annual Winter meeting in Los Angeles in November, 1969.

The error in df/dn due to the algorithm used in the integration scheme has been studied also for the heat-transfer solutions obtained with the improved 1S technique. The fractional error was found to be generally

less than 0.01 for the three cases considered (Kn=0.5, 1, and 10; T_1/T_2=0.7) throughout the velocity space, except that one might expect large errors, e.g., near extrema of the distribution function f. There is a marked decrease in the fractional error when the 1S technique is used.

As indicated above, we have incorporated the Maxwell gas-surface interaction model as the new boundary condition for the heat-transfer problems. Monte-Carlo solutions are being obtained for several sets of parameters. The effect of the gas-surface interaction on the nonequilibrium gas behavior will be studied.

We have made a detailed study of the quadrature errors of calculating the moments of both the distribution function and the collision integral for the heat-transfer problems by comparing with the analytical calculations obtained in our study of the 4-moment heat-transfer solution. It was found that the error in the moments calculated numerically in the Boltzmann program is insignificant for temperature ratios T_1/T_2 larger than 0.3 in the entire range of Knudsen number.

We have also completed (1) a detailed comparative study of a 4-moment and a 5-moment solution of the heat-transfer problems with the corresponding Monte-Carlo solutions of the Boltzmann equation, and (2) an accuracy study of these two-moment solutions by using various tests provided in our program.

4.5. Near-Equilibrium Flows

One of the principal objectives of our research is to study the nonequilibrium behavior of gas-flow problems in which the approach of using the Navier-Stokes equations, valid only for small departure from thermal equilibrium is not applicable. We have applied our method to problems under near-equilibrium as well as far-from-equilibrium flow conditions. Such a success has also made it possible for us to initiate a study of the second-order effect of the Chapmann-Enskog solution on the collision integral. Two parameters used in this study are the local gradients of the temperature and the velocity. That is, we propose to study quantitatively the inaccuracy of the Chapman-Enskog first-order calculation (thus also the Navier-Stokes calculations) as well as the second-order correction

for flows that have large local gradients. Such a study will provide much needed information in studying problems such as hypersonic flow fields in which a large range of property gradient exists. As an example, we have evaluated such errors in a shock wave as functions of position for M = 1.2, 1.56, 2, 2.5, 3, 5 and 10. The second order effect on the density gradient dn/dx for large Mach numbers was found to be most significant near the cold side of the shock wave.

5. HIGH-VOLTAGE BREAKDOWN†

E. M. Lyman
R. Allen
D. Alpert
T. C. Casale
S. Y. Ettinger
T. Matalis
B. Parhami
M. Smith
R. Wallat

5.1. Studies of the Nature of Surface Protrusion which Initiate Breakdown

It has been clearly demonstrated [1] that whiskers or protrusions on the cathode surface of electrodes in ultrahigh vacuum are instrumental in initiating electrical breakdown. As the electrode voltage is increased, field emission from the tip of a protrusion eventually becomes large enough to heat the tip, or the anode spot opposite the emitting tip, to the melting point; a plasma of metallic ions develops which in turn provides the conductive path for the arc or spark.

The nature of the protrusions and the mechanism of their formation has been a subject of considerable interest. Alpert et al. [2] showed that the protrusions could be blunted by gas conditioning, thereby improving the voltage-holding capabilities of the electrodes. Smith and Little [3] observed that the protrusions seem to surround craters on the cathode surface, and that the craters appear to be formed by a charged clump of material impinging at high speed on the surface, melting a small region surrounding the point of impact. The protrusions are metallic fingers drawn out of the molten pool by the strong electric field. They proved that particles inserted into small holes in the anode become charged under the influence of the strong electric field, emerge under the electrostatic forces, and impinge upon the cathode, producing similar craters. Using cylindrical geometry, Chatterton [4] found that clumps of material impinging upon the center cathode wire would, at low voltages simply indent the wire, but, at high voltages, since the energy available is proportional to eV, would produce melted craters, with their accompanying protrusions, and field-emission current where there had previously been none.

Although the experiments described in what follows do not identify precisely the source or the nature of the clumps, they clearly show that some, if not all, of the clumps can be removed from the apparatus by

†Supported principally by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199 and in part by the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

drastic cleaning procedures, thus eliminating the cause of the protrusions. In the absence of protrusions, the maximum electric field before breakdown may be increased by about two orders of magnitude over what has previously been practical, for plane parallel electrodes.

At the outset, the object of our experiments was to study the nature of the protrusions and to determine their origin. At first, the electrodes were flat single-crystal tungsten or copper disks about 10mm in diameter, installed in a metal-and-glass chamber [1]. The electrodes and chamber were cleaned in Alconox and in distilled water, then assembled on a laminar-flow clean bench. The chamber was attached to the vacuum system and baked at 400°C to give a pressure, when cool, of $4 \times 10^{-8} N/m^2$. Despite great care in cleaning, the system obviously was not clean. Microdischarges were always seen during run-up of the voltage; field-emission current was always present for fields of the order of 50MV/m, and the measured field enhancement factor β, was always in the range from 30 to 300, indicating that sharp protrusions had been formed on the surface even before any breakdowns had occurred. A typical crater on a tungsten wire is shown in Fig. 5.1. This large crater, about 50 μm across, is the result of an arc or spark between the electrodes, and shows that multiple strikes have occurred. Protrusions emerge from the edges as well as the interior of the crater. Crystalline patches of various orientations are present in the virgin metal, as well as in the bottom of the crater.

Several attempts were made to produce systems that could be more thoroughly and reliably cleaned, so as to eliminate the clumps of material that were obviously flying around between the electrodes in the previous vacuum chambers. A design that, though cumbersome, has permitted ultrasonic cleaning and induction heating of the electrodes to 1800°C is shown in Fig. 5.2.

The system is all glass except for two electrodes of 0.75-mm-diameter tungsten wire bent as shown to give a well-defined electrode geometry at the crossing region. For a gap-spacing of 0.05mm, the two surfaces at the crossing region may be regarded as approximately plane and parallel, as far as the effects of one electrode surface on the other are concerned. The electrode wires each form a loop to pick up magnetic flux from the induction heater. When the system is evacuated, the contraction

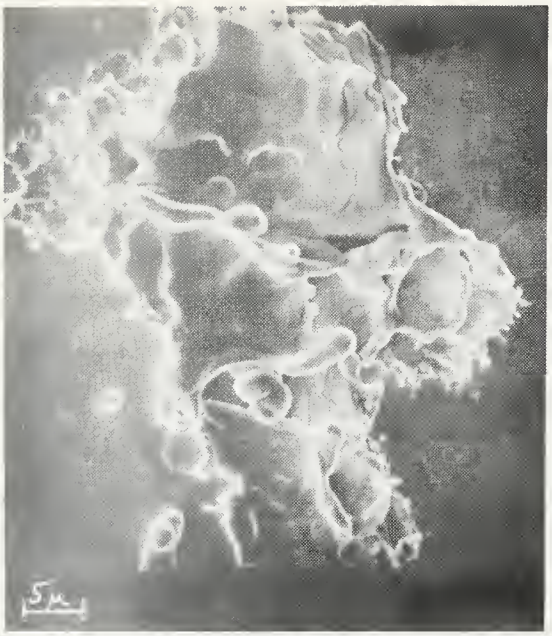


Fig. 5.1. Typical crater and protrusions, after an arc.

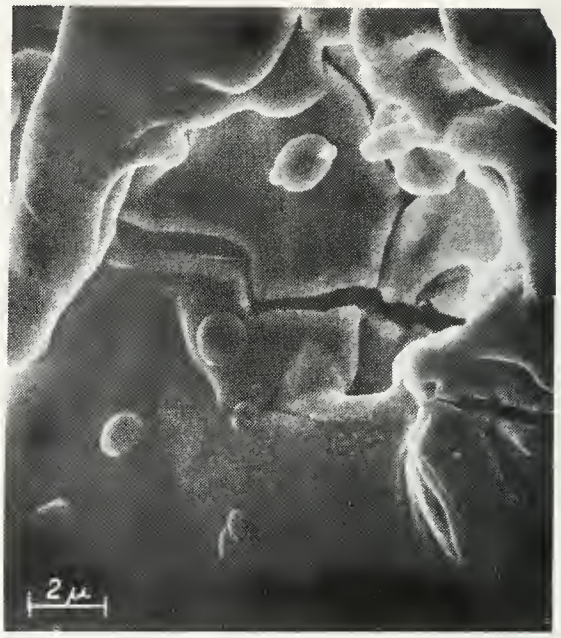


Fig. 5.5. Cracks in the floor of a crater due to metal shrinkage.

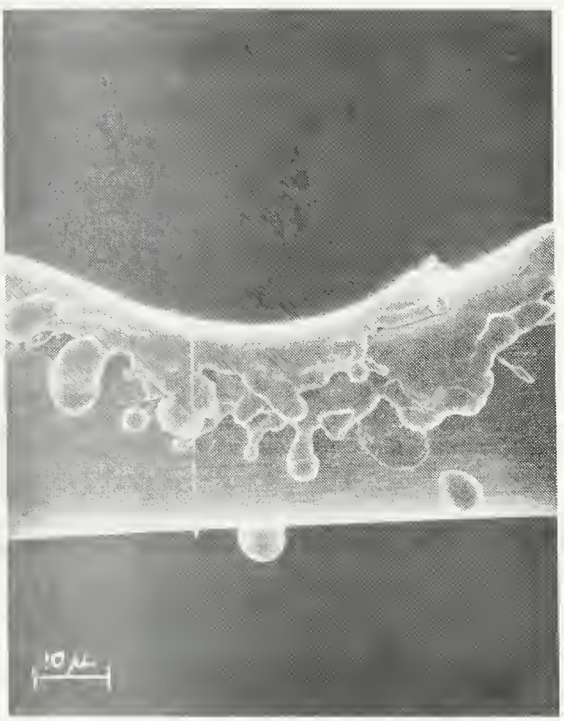


Fig. 5.3. One of a pair of electrodes welded together by a high-voltage breakdown.

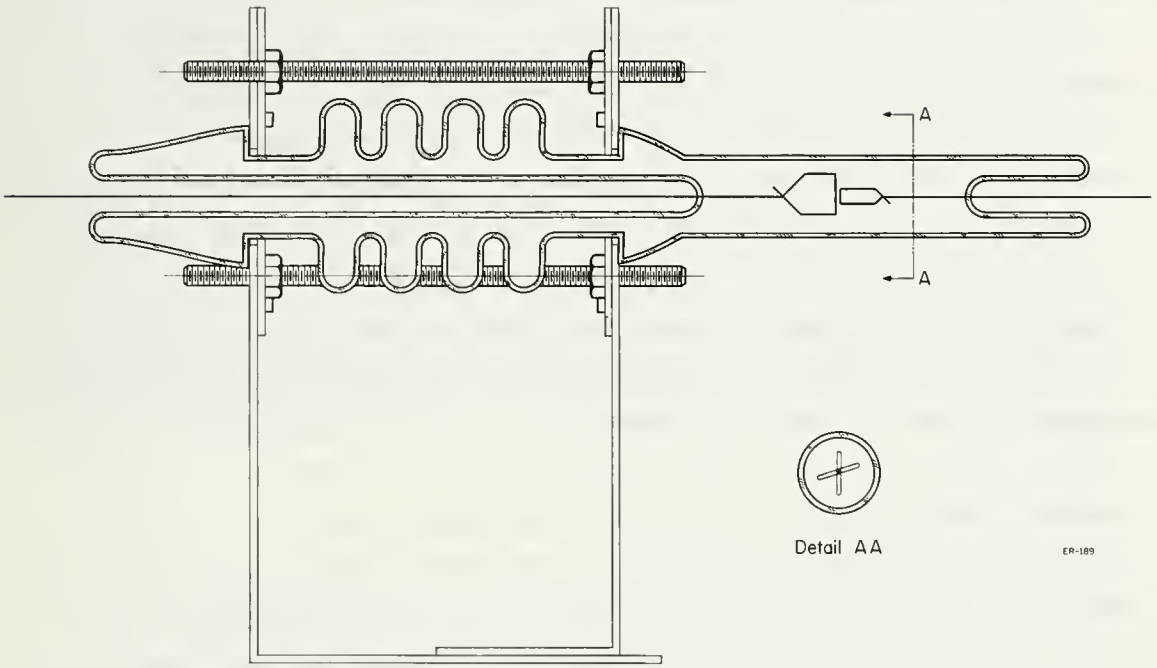


Fig. 5.2. Vacuum chamber. The wire loops cross at approximately right angles to give well-defined electrode geometry of the facing electrodes and to facilitate induction heating them individually.

of the glass bellows would allow the electrodes to touch (a disaster) if it were not for the external adjustable spacers. The bellows permit a total gap adjustment of 0.5mm, which is enough so that the electrodes can be prevented from touching when induction heated, one at a time, to 1800°C.

A typical procedure for preparing the electrodes and chamber is as follows: Straight, commercial, 1-mm, ground, tungsten rod is electropolished in a solution of 1 g of NaOH to 100 mℓ of distilled water. This smooths off the wire, (although the Scanning Electron Microscope reveals that the wire still has some protrusion and drawing marks on it; in fact, the commercial-grade tungsten appears as if composed of many cylindrical strands.) After electropolishing for 10 minutes, the wire could usually be bent without breaking into a loop about 10mm in diameter as shown. After the wires are sealed into the glass presses, they are mechanically polished with 600-grit carborundum paper and AlO_2 to eliminate large nodules of glass or WO_2, and then electropolished again down to 0.7mm diameter. The presses are then sealed into the glass tube containing the bellows, leaving a gap of about 0.20 mm between the wires prior to evacuation. A cleaning technique, arrived at without much experimentation, consists of a series of 5-minute treatments in the ultrasonic cleaner, two with Alconox solution, four each with distilled water, and then double-distilled water. Great care is taken to see that the liquid completely fills the chamber each time.

After cleaning, the tube is glassed on to the ultrahigh-vacuum system, evacuated (after attaching the spacers), and baked for 5 or 10 hours at 400°C after which the vacuum (when the system is cool) is about $1 \times 10^{-8} N/m^2$ as read on the Bayard-Alpert gauge.

Although the above design is very simple, it was arrived at somewhat laboriously by trial and error. For example, it was found that spot welding usually failed under prolonged ultrasonic cleaning, so that electrodes consisting of high-purity tungsten wire spot welded to tungsten stanchions were likely to break during cleaning. Those that survived ultrasonic cleaning failed in another way when used as electrodes; the electrostatic forces on the fine wires would pull the electrodes together at high fields. A typical example of one of a pair of electrodes which

welded together when the fields reached about 10×10^8 V/m is shown in Fig. 5.3. Drawing marks are much more pronounced than usual on this piece of high-purity tungsten wire 0.50 mm in diameter.

An optical gap-monitoring arrangement was necessary for measuring the electrode surfaces spacing which was found to decrease by about 25% at high voltages due to electrostatic forces on the electrodes and on the glass envelope.

In a typical experiment for measuring the field-emission current and breakdown voltage, the anode was grounded through a 100-kΩ resistor, and high voltage was applied to the cathode. Microdischarges, if any, were observed on an oscilloscope as voltage pulses across the 100-kΩ resistor. Field emission was measured by a series meter. It was generally observed that if the electrodes had been induction heated for 10 minutes each at 1750°C, there were no microdischarges and the field emission current was less than 1×10^{-10} A, as the voltage was raised. A violent breakdown usually occurred at a field of between 10 to 20×10^8 V/m in the gap. On the other hand, if the electrodes had not been heated to 1750°C, microdischarges always occurred as the voltage was raised. At moderately high fields, a microdischarge sometimes resulted in subsequent, steady, field-emission current indicating that a crater with protrusions had been formed. As the voltage was raised breakdown would usually occur at a field of less 5×10^8 V/m. The gap spacing was necessarily small (0.04 mm) when the highest fields were achieved, because the maximum electrode voltage that could be applied was limited by external sparking and corona to about 60 kV.

In some cases it was possible, after breakdown or cratering, to restore the cathode to its initial smooth and nonemitting state by heating it to 1750°C for a few minutes.

In order to view the electrodes in the Scanning Electron Microscope (SEM), it was unfortunately necessary to cut the tube apart, and to mount the facing portions of the two electrodes on small specimen holders 12 mm in diameter. A specimen holder may be rotated, translated and tilted in the SEM, thus producing "views" from various angles. In our stereoscopic pairs, the tilt angle was changed in the two pictures by $\Delta\theta = 7^\circ$ to 20°, depending on the magnification. Figures 5.4a and 5.4b are a stereo pair of the portion of a large crater shown in silhouette in Fig. 5.4c.

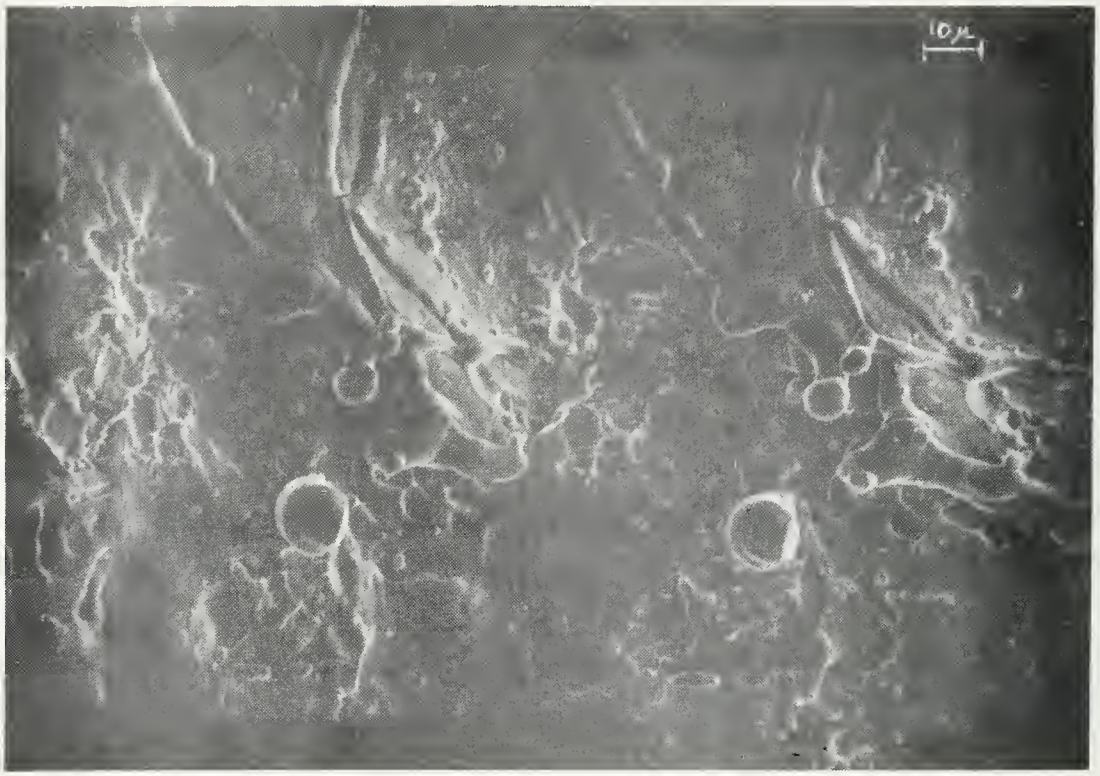


Fig. 5.4a, b. Stereo photographs of a crater after breakdown.

Fig. 5.4c. Profile view of the crater of Figs. 5.4a and 5.4b.

The instrumental resolution is about 50 nm. The picture gives the impression that some large-diameter protrusions were formed and that the tips of these protrusions suffered further melting and rounding off during the arc. The enhancement factor obtained from the Fowler-Nordheim plot, after breakdown, is $\beta = 23$. A search of this picture fails to reveal any protrusions sharp enough to give, from the geometry, a value of β greater than 11. The large dome gives $\beta \approx 4$. It is presumed that the protrusion responsible for $\beta = 23$ is located elsewhere in the crater, out of the field of view of this picture or that it is too small to be seen readily at this magnification. (In other experiments emitting protrusions have been found to be about 1 μm in height and 0.1 μm in diameter.)

Cracks in the floor of the crater due to cooling are visible in Fig. 5.4, and a more pronounced example is shown at higher magnification in Fig. 5.5.

5.2. Conclusions

In a "clean" system in ultrahigh vacuum, electrical breakdown is initiated as the result of field emission from the sharp tips of protrusions on the cathode. Initiation occurs when either the tip of the protrusion, or the spot on the anode opposite it, is heated to the melting point by the field-emission current. The molten surface becomes distorted and vaporizes, the material becomes ionized in the electron stream, and the resultant plasma of metallic ions provides the conducting path for the spark or arc.

The presence of charged clumps that impinge upon the cathode surface is signalled by microdischarges. The production of a protrusion is signalled by a microdischarge followed by a steady field-emission current where previously there had been none. By removing the clumps so that there are no microdischarges, it has been possible to obtain average fields of from 10 to 20×10^8 V/m in the gap, without field emission or breakdown, a factor of 10 to 100 better than the previous limit.

5.3. References

1. D. Alpert, D. A. Lee, E. M. Lyman and H. E. Tomaschke, J. Vac. Sci. Technol. 1 35, (1964).

2. D. Alpert, D. Lee, E. M. Lyman and H. E. Tomaschke, J. Appl. Phys. 38, 880 (1967).

3. R. P. Little and S. T. Smith, "Investigations into the source of sharp protrusions which appear on flat cathode surfaces as a result of the applications of high electric fields", pp. 41-49, Proc. Second Int. Symp. on Insulation of High Voltages in Vacuum, MIT, Sept. 1964.

4. P. I. Biradar and P. A. Chatterton, "Cathode Surface Structure and Prebreakdown Currents," pp. 35-40. Proc. Third Int. Symp. on Discharges and Electrical Insulation in Vacuum. Paris, Sept., 1968.

6. SPACE AND ATMOSPHERIC SCIENCES

A. Cohen
D. H. Cooper
G.R. Karr
H.W. Knoebel
J.L. Myers
D. Skaperdas
S.M. Yen

6.1. Status of Feasibility Study for the Relativity Satellite[†]

As previously reported [1], the Laboratory's study of the feasibility of testing the Einstein theory of gravitation by means of a satellite experiment using a passive, unshielded gyroscope had been essentially completed with the issuance of a thesis report [2] of an optimal design whose possibility had been indicated in the earlier final report [3]. During the present reporting period, the overall technical status of this feasibility study has been under review, and a summary statement of this status has been prepared. The making of such a statement is felt to be appropriate at this time, because it is believed that these studies have reached a new end point in an affirmative demonstration of feasibility. No further studies of a feasibility nature are thought to be profitable or necessary. Indeed, the logical next step would be the actual design and conduct of the experiment itself.

These studies have concluded that an optimal design for the experiment, predicating the use of the existing Baker-Nunn observing facilities of the Smithsonian Astrophysical Observatory, would have the following characteristics:

1. An overall probable error for measuring the Einstein gyroscopic precession (6.7 arcsec per year) of 0.3% for a two-year run based on an observing efficiency of 65%, the percent of theoretically-possible observing opportunities actually exploited because of losses due to weather, etc.
2. Broad tolerances for the optimal parameters of the orbit -- altitude (580 km, plus 400 km or minus 200 km), inclination ($31^{\circ} \pm 5^{\circ}$), and spin-axis orientation (essentially equinoctal, $\pm 1^{\circ}$ out of the equatorial plane and $+5^{\circ}$ in the plane).
3. High optical reliability -- the optimal orbit carries the satellite within observing range of every one of the Baker-Nunn cameras, so that the 6 mirrors provide spin-axis data via solar flashes of

[†]Supported in part (through June 30, 1967) by the National Aeronautics and Space Administration under Research Grant NsG-443 and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-0199.

an optimal brightness that is at least five-fold greater than minimum useful values, and so that the spherical surface provides orbital data via a continuous background brightness whose photography requires only a 2% tracking accuracy from the Baker-Nunn cameras.

4. Data-reduction procedures which yield a high tolerance for predictable background precessions -- the observed motion is analyzed into secular and several periodic terms representing differing predictable combinations of gravity-gradient, radiation-pressure, aerodynamic, and relativistic effects, predictable as to form, with coefficients to be determined by fitting to the observed data.

The optimal orbital parameters were determined by exercising a computer-based simulation of the conduct of essentially the whole experiment, using realistic observation errors and the data-reduction procedure described. By varying the parameters, an overall view of the experiment could be obtained, with remarkable results. For example, it was possible to see that previous efforts to minimize gravity-gradient precessions by adjusting the orientation of the spin axis in orbits of essentially zero inclination not only constrained the experiment to excessively tight specifications in spin-axis orientation, but also caused an increase in overall error through lost observations from the higher-latitude stations. An increase in inclination increases the data rate and causes a part of such secular background precessions to appear as periodic, whereby a means for their independent estimation is provided. What is important, since the background precessions are of no relativistic interest, are the consequences that the contribution from the relativity precession is then more accurately measured, and that moderate departures in spin-axis orientation from the optimal values would have only a slight effect. The optimization in all orbital parameters was then carried forward with reference to the error in the relativity measurement.

This summary is too brief to allow a description of all the benefits of the optimal-design approach, of which the slight dependence of a performance index upon the exact choice of parameter values is only one. Though the optimal choices in the final design may differ somewhat from those quoted -- a nonuniform distribution of expected weather losses from station to station

could alter the optimal inclination to a slight extent, for example -- reasonable errors in knowing the optimal values and in achieving them will not be serious. Even a further loss in half the observations could be compensated by extending the run by 25%. Again, the ability to exploit the predictability of the gravity-gradient precession, so that comparatively large amounts may be tolerated, permits a design of the gyro's inertial figure that reduces the random effects of micrometeoric impingement (except for rare large events which may be accorded special treatment) to fall well within the probable error. Perhaps the greatest benefit is the strengthened confidence that is to be obtained from an integrated view of the design of the whole experiment.

The Laboratory's view of the experiment has undergone considerable evolution during the course of the study. Only within a few months of the end of the original funding period for the study did it become clear that the overall accuracy of the experiment could be improved from the initial goal of 10% to achieve 2% through an increase in diameter of the gyro, a reduction in the number of mirrors to 6 in nonredundant locations, and making suitable accommodations in other parameters, as reported in Ref.[3].

The study was continued at a reduced level after the termination of the funding period in order to further confirm the suitability of these accomodations. It was during this period that the large-scale simulation of the experiment provided the new approach to optimal design and showed the possibility of further improvements in accuracy. A full account of this subsequent work appears in Ref.[2]. This represents the view of the experiment upon which we have ultimately stabilized, and it predicts the overall probable error of 0.3% for the measurement of the relativistic precession.

The evolution of this view has been so substantial that it is almost as if a new experiment were being considered, though the overall plan is, in essence, the same. The new accuracy figure certainly places it in a new position of relevance to Einstein's theory and the challenges that theory currently faces in accounting for the consequences of the recent measurement [4] of the oblate figure for the sun. This oblateness would augment the classical accounting for the perihelion precession of the orbit of Mercury to leave a relativistic residue that is some 8% smaller than the Einstein value, a value that hitherto had been thought to have been confirmed with an accuracy of 1%. The possibility of making an independent check of the theory with an

accuracy of 0.3% is obviously of major scientific importance in this circumstance. Such would be especially true of an experiment that, like the orbit of Mercury, provides a challenge to the deeper parts of the theory than its null-geodesic aspects. The gyroscopic test is such a deeply-challenging one, as Schiff has observed [5]. One may expect that it could discriminate among the theoretical adjustments [6,7] that may be offered to account for the apparent 8% defect in Mercury's relativistic motion.

Though possibly of secondary urgency in terms of immediate scientific interest, the measurement of the Lense-Thirring component in the relativistic gyro precession [5,8] becomes possible because of the improved accuracy of the experiment. This component, caused by the rotation of the Earth, is about 1.5% of the total relativistic precession, and it would augment the total for retrograde orbits, otherwise diminishing it. A full confirmation could be obtained by a second launch using a retrograde orbit after a successful demonstration of the performance of the first experiment. In such a way, the Lense-Thirring effect could be measured to within 15% of its value.

Regarding the conduct of the experiment, conversations with the Smithsonian Astrophysical Observatory have indicated that the observing burden would fall in the range of some 10 to 30 percent of that normally carried by them. Pending confirmation of the experiment's feasibility and scientific importance, there was an expression of interest and willingness for making the necessary observations. Also, discussions with Owens-Illinois have indicated a capability for the complete fabrication of the gyro-satellite body using their Cer-Vit material.

Thus, it is believed that all points of feasibility, including many that could not be mentioned in such a brief summary, have been covered in the study. The status is that of a demonstrated feasibility for an experimental accuracy at least 30-fold better than had been initially anticipated, for an experimental design making for high reliability and low cost because of its tolerance of moderate parameter maladjustments and its technological simplicity, and for an experiment of eminent scientific importance.

D.H. Cooper
H.W. Knoebel

6.2. Study of an Orbiting Gyro Experiment to Measure Gas Density and Gas-Surface-Interactions Parameters[†]

6.2.1. Review of the Objective of the Present Study

The objective of the present study is to carry out analytical and design studies for a satellite experiment, using a single, passive, spinning gyro, to determine the absolute orbital gas density and two gas-surface-interaction parameters. These parameters are to be determined from measurements of the drag, the spin-slow-down rate, and the spin-axis-precession rate. Measurements would be made using techniques developed at the Coordinated Science Laboratory for a similar satellite system. (See 6.1, above).

G.R. Karr

6.2.2. Program of the Proposed Optimization Study

A program formulated for making the optimization study consists of five phases. The tasks to be performed in each phase and their significances are described as follows:

6.2.2.1. Phase One

Phase One is to develop the aerodynamic equations for a spinning satellite. The aerodynamic behavior of spinning satellites, more so than nonspinning satellites, is strongly dependent upon the choice of the various proposed gas-surface-interaction models. To provide maximum flexibility in the study of the effects of using the various proposed models to describe the molecular reflection, a generalized gas-surface-interaction model was developed which contains the current models as a subclass and allows the possible inclusion of more sophisticated models which may yet be proposed. The aerodynamic equations for spinning bodies were then developed with the incorporation of the proposed generalized gas-surface-interaction model. These equations can, in general, be applied to any satellite which contains no concave surfaces.

6.2.2.2. Phase Two

Phase Two is to develop numerical techniques and computer programs

[†]Supported in part by the National Aeronautics and Space Administration under Research Contract NAS 8-21442.

for evaluating the aerodynamic properties of satellites of selected shapes and to make preliminary optimization studies. Two basic computer programs will be developed in this phase.

a. One computer program will be to determine the instantaneous aerodynamic properties of a spinning satellite as a function of the angle of attack and the gas-surface-interaction parameters.

b. The other computer program will be to determine the precession rate, spin-slow-down rate, and drag effects as a function of the orientation of the satellite in the orbit and the gas-surface-interaction parameters.

These two computer programs will form the basis for making the optimization study.

An initial optimization study will be made with several assumptions to facilitate testing the basic computer programs and to obtain detailed preliminary results.

The first computer program will be used to study the aerodynamic properties of the satellite as influenced by the gas-surface-interaction parameters and the satellite-system parameters, such as the angle of attack and the spin rate. These studies will be made for satellites of appropriate shapes so as to obtain results useful for the studies in the next part of the study.

The results of the study described above will be used as the necessary input data in the second computer program for studying the drag, the spin-slow-down rate and the spin-axis-precession rate as influenced by the orbital parameters and parameters pertaining to the orientation of the satellite.

Methods developed and some of the results obtained in this phase would, themselves, be significant in solving other problems of satellite aerodynamics as well as analyzing the data from other satellites. The results obtained would also be of general interest since the aerodynamics of spinning satellites have not been well understood and the effects of different gas-surface interaction models in the aerodynamic analysis of satellites have not been studied in a systematic manner.

6.2.2.3. Phase Three

Phase Three is to determine the relative accuracy with which the three parameters of interest can be measured with respect to the satellite

shape, orientation of satellite in orbit, and the gas-surface-interaction parameters. These determinations will be made first for the assumptions of constant atmospheric density, circular orbits, and hyperthermal flow.

Upon completion of this phase of the study, preliminary conclusions may be drawn concerning the selection of the satellite shape and the importance of the gas-surface-interaction models. These are the major factors which influence the optimization and general feasibility of the experiment.

Since the accuracy of making drag measurements is dependent on the eccentricity of the orbit, this study will be refined by considering the cases of noncircular orbits and the use of an elementary atmospheric model.

Before final selection of the satellite system parameters are made, the perturbing effects of the orbital environment must be included in the optimization study. The perturbing effects are divided into two areas of study (aerodynamic perturbations and general space-environmental perturbations) to be made in Phase Four and Phase Five respectively.

6.2.2.4. Phase Four

Phase Four is to extend the aerodynamic analysis to include the thermal motion of incident gas molecules and currently accepted atmospheric models. This phase will also include a study of the effects on the experiment of using orbits of high eccentricity when the aerodynamic effects are maximum during perigee. The study of additional aerodynamic factors that would influence the final selection of satellite shape and satellite orientation will be completed in this phase.

6.2.2.5. Phase Five

Phase Five is to extend the optimization study by including the effects of solar radiation pressure, gravity gradient, meteorite bombardment, surface gas absorption, electrical-magnetic interaction and other effects which are found to influence the optimization of the satellite system. These effects will primarily influence the maximum possible accuracy of the proposed experiment.

G.R. Karr

6.2.3. Summary of Progress on Optimization Study

A qualitative study was performed on a typical composite-shape satellite to illustrate the feasibility of the proposed experiment. This

analysis also indicates a possible procedure for determining an optimum satellite shape.

6.2.3.1. The Satellite Shape

The cone-sphere composite satellite shape is illustrated in Fig. 6.1. The conical segment is attached to the spherical segment such that the two parts have the same surface slope at the line of joining. The spin axis is along the axis of the cone and, for this preliminary analysis, the centroid of mass is placed at the center of the sphere defined by the spherical segment. This design has the advantage of containing no concave surfaces that could allow molecules reflecting from one part of the surface to again impinge on the satellite surface. The principal criterion for choosing a satellite shape is that the dynamical equations representing the drag, slow down, and precession of the satellite be independent in the three parameters of interest. This criterion is met in the cone-sphere design due to the different aerodynamics of the spherical segment and the conical segment at an angle of attack.

6.2.3.2. Qualitative Description of Torque Characteristics

Consider the cone-sphere composite satellite orientated such that the spin-axis is out of the orbital plane at an angle equal to the half angle of cone. (See Fig. 6.2) At one point in the orbit, the satellite is orientated with respect to the molecular flow such that molecules impinge on the conical segment when the axis of the cone is at an angle of attack. The satellite will then experience a torque, perpendicular to the spin axis, at this point in the orbit. At 180° later in the orbit, the molecular flow will be reversed and, therefore, will impinge on only the spherical segment of the satellite. (See Fig. 6.3) Since the centroid of mass has been placed in the center of the spherical segment, the satellite experiences zero torque at this point in the orbit. The torque experienced by the satellite at the $\pm 90^{\circ}$ positions in the orbit cancel due to symmetry. The total effect is thus a net torque acting perpendicular to the spin axis averaged over each orbit of the satellite.

6.2.3.3. Preliminary Analysis of the Composite-Satellite Dynamics

The forces on the satellite are determined using the Maxwell gas-surface reflection law with α_d equal to the fraction of molecules reflected diffusely, the remaining fraction is reflected specularly, and α_T equal to the amount of thermal accomodation of the diffusely reflected fraction of molecules. For a satellite of the above description and orientation, the dynamical equations

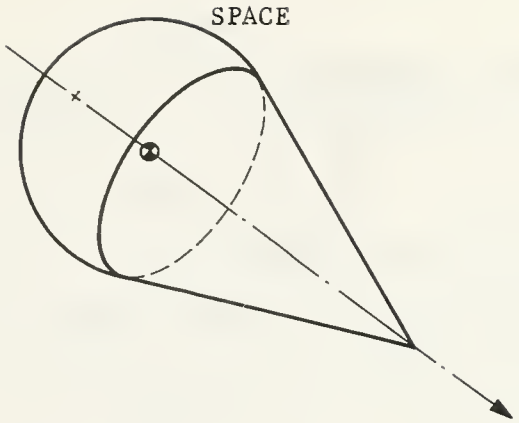


Fig. 6.1. Cone-sphere composite shape.

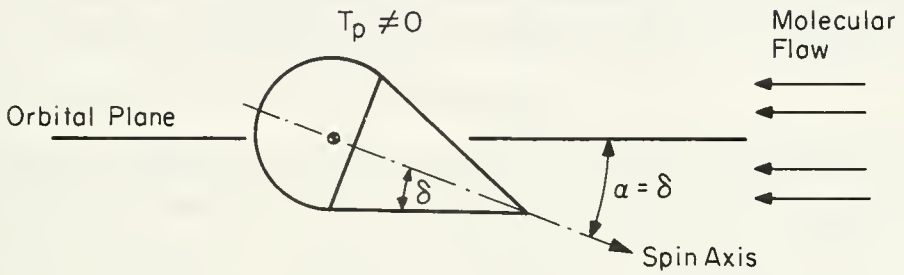


Fig. 6.2. Orientation of satellite at 0° position in orbit.

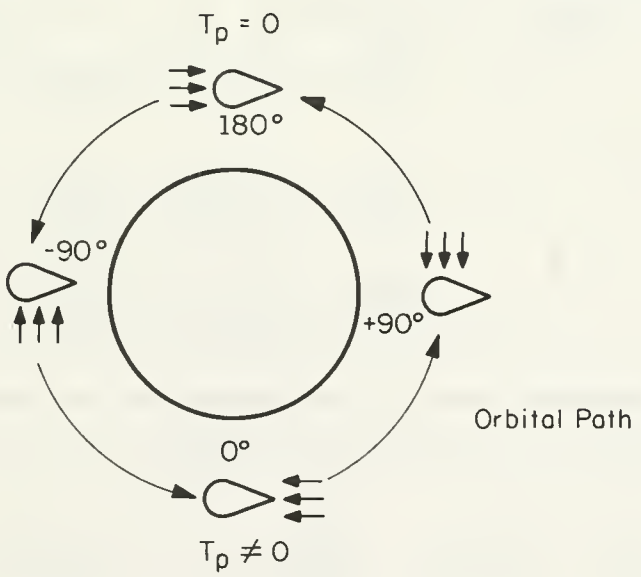


Fig. 6.3. Illustration of satellite orientation with respect to molecular flow during one orbital pass.

RR-407

are in general of the following form:

Drag:

$$D = A_1\rho + A_2\rho\alpha_d + A_3\rho\alpha_d\ (1-\alpha_T)^{\frac{1}{2}},$$

Torque perpendicular to spin axis:

$$T_p = B_1\rho + B_2\rho\alpha_d + B_3\rho\alpha_d\ (1-\alpha_T)^{\frac{1}{2}}\ ,$$

Slow-down torque:

$$T_s = C\rho\alpha_d\ ,$$

in which the coefficients $A_{1,2,3}$, $B_{1,2,3}$ and C are known constants dependent upon the shape, angle of attack, size, and density of the satellite. A solution exists for the parameters ρ, α_d and α_T provided that

$$A_1B_3 - A_3B_1 \neq 0.$$

The relative accuracy in which the three unknown parameters can be determined can be assesed by performing an error analysis on the above relations. For the purposes of this preliminary analysis, assume that the coefficients $A_{1,2,3}$, $B_{1,2,3}$ and C are known so that error in determining ρ, α_d, and α_T will depend upon the accuracy of measuring the drag, precession and slow down torque. This dependence can be represented as follows

$$\Delta\rho = \frac{\partial\rho}{\partial D}\Delta D + \frac{\partial\rho}{\partial T_p}\Delta T_p + \frac{\partial\rho}{\partial T_s}\Delta T_s,$$

$$\Delta\alpha_d = \frac{\partial\alpha_d}{\partial D}\Delta D + \frac{\partial\alpha_d}{\partial T_p}\Delta T_p + \frac{\partial\alpha_d}{\partial T_x}\Delta T_s,$$

$$\Delta\alpha_T = \frac{\partial\alpha_T}{\partial D}\Delta D + \frac{\partial\alpha_T}{\partial T_p}\Delta T_p + \frac{\partial\alpha_T}{\partial T_s}\Delta T_s,$$

where the difference symbol, Δ, represents the uncertainty.

When the derivatives are evaluated, the following equations are obtained:

$$\frac{\Delta\rho}{\rho} = \frac{-B_3D}{\rho E}\left(\frac{\Delta D}{D}\right) + \frac{A_3T_p}{\rho E}\left(\frac{\Delta T_p}{T_p}\right) + \frac{A_2B_3-A_3B_2}{E}\alpha_d\left(\frac{\Delta T_s}{T_s}\right)$$

$$\frac{\Delta\alpha_d}{\alpha_d} = \frac{-B_3D}{\rho E}\left(\frac{\Delta D}{D}\right) + \frac{A_3T_p}{\rho E}\left(\frac{\Delta T_p}{T_p}\right) + \left(1 + \frac{A_2B_3-A_3B_2}{E}\alpha_d\right)\frac{\Delta T_s}{T_s}$$

$$\frac{\Delta\alpha_T}{\sqrt{1-\alpha_T}} = \frac{2B_1D}{\rho E}\left(\frac{\Delta D}{D}\right) - \frac{2A_1T_p}{T_p}\left(\frac{\Delta T_p}{T_p}\right) + \left(2\alpha_d\frac{A_1B_2-B_1A_2}{E} + 2\right)\frac{\Delta T_s}{T_s},$$

where $E = A_1B_3-A_3B_1$.

The above equations illustrate a possible procedure for determining the optimum shape of satellite for measuring the three parameters accurately. As mentioned earlier, the coefficients A, B, and C depend upon the angle of attack of the satellite, the shape parameters and other quantities. The procedure then is to find the nominal values for these quantities which will yield the least error in determining α_d, α_T, and ρ.

A two-dimensional analysis of the cone-sphere satellite shape was made to obtain approximate values for the torques and drag of such a satellite. For a 60^o cone angle, the following results were obtained.

$$\left|\frac{\Delta\rho}{\rho}\right| < 3\left|\frac{\Delta D}{D}\right| + 1.75\left|\frac{\Delta T_p}{T_p}\right| + 0.5\left|\frac{\Delta T_s}{T_s}\right|,$$

$$\left|\frac{\Delta\alpha_d}{\alpha_d}\right| < 3\left|\frac{\Delta D}{D}\right| + 1.75\left|\frac{\Delta T_p}{T_p}\right| + 1.5\left|\frac{\Delta T_s}{T_s}\right|,$$

$$\left|\frac{\Delta\alpha_T}{\sqrt{1-\alpha_T}}\right| < 7.5\left|\frac{\Delta D}{D}\right| + 7.5\left|\frac{\Delta T_p}{T_p}\right| + 3\left|\frac{\Delta T_s}{T_s}\right|.$$

The above results represent only an upper bound on the error expected due to inaccuracies in measuring D, T_p, and T_s.

6.2.3.4. Gas-Surface Interaction Models

Numerous models have been proposed to represent the interaction of neutral molecules with surfaces moving at satellite velocities. Due to the lack of experimental data, none of these models have been substantiated as to its applicability to satellite aerodynamic analysis.

In the preliminary feasibility study, the Maxwell model has been used in the analysis of the aerodynamics of the proposed satellite. This model requires only two parameters to represent the interaction completely: α_d, which is the fraction of the molecules reflected diffusely, the remaining fraction is reflected specularly; and α_T, which represents the amount of thermal accomodation or energy transfer of the diffusely reflected fraction of the molecules. Since the proposed satellite would have the capability of

determining these two parameters accurately, complete determination of the gas surface interaction would result by using this model in the aerodynamic analysis.

On the basis of the interpretation of recent laboratory results, there has been a general lack of acceptance of the Maxwell model in the description of the molecular reflection. For this reason, models other than the Maxwell model are being studied as to their applicability to the aerodynamic analysis of the proposed satellite.

R. Schamberg [9] has proposed a model which has three parameters: a thermal accomodation coefficient and two parameters which determine the direction and diffusiveness of the reflection. The latter, found to have only a small influence on the measurable effects of the interaction, can be eliminated. A two-parameter Schamberg model could then be determined by the proposed satellite.

Another model which can be used in a two parameter representation is that of Nocellia [10]. In this model, the reflection is represented by a half Maxwellian distribution having a mass velocity of undetermined direction and magnitude.

Also to be included in this list of models are the many reflection patterns suggested by laboratory experimental results and computer-simulation techniques.

Use of as many of these models as possible will be made in the analysis of the dynamics of the proposed satellite. K. Moe [11] has used a similar technique in the analysis of a paddlewheel satellite. By using this technique and comparing results with the dynamics of satellites of other shapes, it is felt that one model may give consistent results and thereby determine the accuracy of the proposed models. If no presently proposed model gives this consistency, a procedure could be developed to propose a new model which would better represent the actual gas-surface interaction on satellite surfaces.

6.2.3.5. Generalized Gas-Surface-Interaction Model

In order to be able to incorporate various gas-surface-interaction models in the aerodynamic analysis of spinning satellites, a generalized gas-surface-interaction model was developed.

The generalized reflection model is basically an extension of Schamberg's original model. The result, however, is a model which, by proper

choice of parameters, can be reduced to represent not only Schamberg's original model, but also Maxwell's model, Nocilla's model, and models suggested by recent laboratory experiments. The generalized model is particularly suited for the aerodynamic analysis of the proposed satellite in which the effect of different models on the satellite dynamics is desired. In the proposed model, the incident gas flow will be represented as an impingement at an angle θ, measured from the surface, at a velocity U, relative to the surface. The reflection is represented by a reflected velocity, U_j, and an angle of reflection of θ_j measured from the surface. The parameters U_j and θ_j are representative values of the reflection and the j subscript notation is allowed to take on values of 1,2,3...to provide for reflection models which have more than one reflection component. For example, the Maxwell model has a diffuse and specular component of reflection which would be represented by U_1 and U_2 with $U_1=U$, $\theta_1=\theta$, (specular component) and U_2 the diffuse velocity with $\theta_2 = 90^\circ$.

The magnitude of the reflected velocity, U_j, is related to the velocity of impingement, U, in the conventional manner through the thermal accommodation coefficient

$$U_j/U = \left[1+\alpha_j(T_w/T_i-1)\right]^{\frac{1}{2}} \approx \left[1-\alpha_j\right]^{\frac{1}{2}},$$

where T_w and T_i are the satellite surface temperature and incident gas temperature respectively. Each j component of reflection has a corresponding thermal accommodation coefficient, α_j.

At least one gas-surface interaction model, the Schamberg model, provides for a relationship between the angle of incidence and angle of reflection. In Schamberg's model this relationship is given as

$$\cos\theta_j = (\cos\theta)^{\nu} \qquad \nu \geq 1,$$

where, when $\nu = 1$, the reflection is specular, and when $\nu \to \infty$, then $\theta_j \to 90^\circ$, which represents diffuse type reflection. The Schamberg model is limited to angles of reflection such that

$$\theta \leq \theta_j \leq 90^\circ.$$

However, there is experimental evidence to suggest that in some cases θ_j can be less than θ, which is an overspecular reflection. Also θ_j can be greater than 90° which is sometimes described as backscatter. Since Schamberg's

original model will not provide for overspecular and backscatter type reflections, a new angular relationship was developed to provide more generality to the model and therefore allow incorporation of recent experimental results in the aerodynamic analysis of the proposed satellite.

In developing a relationship between θ and θ_j, a few requirements must be met. First, for every θ there must be one and only one value θ_j. Second, when $\theta = 90^o$, which means that the gas flow is normal to the surface, θ_j must also be 90^o. Third, the obvious requirement that $0 \leq \theta_j \leq 180^o$. With these physical restrictions on the reflection, numerous functions, $\theta_j = f(\theta)$, can be generated to satisfy the general requirements. Since there is little or no experimental evidence to advance a functional relationship of any particular type, the relationship to be chosen should be convenient to work with and general enough to be a first approximation to what ever the relationship might really be. These requirements suggest a linear relationship between θ_j and θ of the form

$$\theta_j = a_j + b_j\theta,$$

where a_j and b_j are constants. The above linear relationship is satisfactory for $\theta \leq \theta_j \leq 180 - \theta$, but violates the physical requirements when an overspecular reflection is to be represented. Therefore, to allow for overspecular reflections in the model and also to add more generality to the representation, a discontinuous linear relationship is chosen in which

$$\theta_j = a_j + b_j\theta, \text{ for } \theta \leq \theta^*,$$

$$= c_j + d_j\theta, \text{ for } \theta \geq \theta^*,$$

where a, b, c, d and θ^* are constants which can be considered separate parameters or fixed values. The constants a, b, c, d, and θ^* can also be written as functions of only one parameter.

Finally, each j component of reflection has a mass flux equal to σ_j, a fraction of the incoming mass flux. The requirements for σ_j are

$$0 \leq \sigma_j \leq 1 \quad \sum_{j=1}^{n} \sigma_j = 1,$$

where n is the number of reflection components.

An additional parameter may also be incorporated, if desired, that would define the beam width of each component of reflection. Schamberg uses

the notation ϕ_o to represent the beam width parameter. However, Schamberg found ϕ_o to have little effect in the aerodynamic analysis of satellites.

6.2.3.6 Aerodynamic Equations

The forces on an element of surface of a spinning satellite can be expressed as

$$d\underset{\sim}{F} = -(\underset{\sim}{U} - \sigma_j \underset{\sim}{U}_j)\rho \underset{\sim}{U} \cdot \underset{\sim}{n} dA, \tag{1}$$

where $\underset{\sim}{U}$ is the velocity of the free-stream gas with respect to the surface, i.e.,

$$\underset{\sim}{U} = \underset{\sim}{U}_\infty - \underset{\sim}{\Omega} \times \underset{\sim}{R}, \tag{2}$$

where $\underset{\sim}{U}_\infty$ is the free-stream gas velocity with respect to the center of mass of the satellite, $\underset{\sim}{\Omega}$ is the angular velocity of the element of surface at the radial position $\underset{\sim}{R}$ from the center of mass. The vector $\underset{\sim}{U}_j$ in equation (1) is the reflected velocity vector with respect to the surface, $\underset{\sim}{n}$ is the surface normal, ρ is the free stream gas density, dA is the surface element area and σ_j, which insures conservation of mass at the surface, is treated as a parameter.

In a similar manner, the torque about the center of mass of the satellite can be expressed as

$$d\underset{\sim}{L} = \underset{\sim}{R} \times (\underset{\sim}{U} - \sigma_j \underset{\sim}{U}_j)\rho \underset{\sim}{U} \cdot \underset{\sim}{n} dA \tag{3}$$

The reflected velocity, $\underset{\sim}{U}_j$, can be expressed as a function of $\underset{\sim}{U}$ by the use of the thermal accommodation coefficient, α_j as follows:

$$\underset{\sim}{U}_j = \sqrt{1-\alpha_j}\ \{f_1[\underset{\sim}{U} \cdot \underset{\sim}{\Omega} \times \underset{\sim}{R}]\underset{\sim}{\Omega} \times \underset{\sim}{R} + f_1[\underset{\sim}{U}_\infty \cdot \underset{\sim}{n} \times (\underset{\sim}{\Omega} \times \underset{\sim}{R})]\underset{\sim}{n} \times (\underset{\sim}{\Omega} \times \underset{\sim}{R}) + f_2\, U\, \underset{\sim}{n}\}, \tag{4}$$

where f_1 and f_2 are functions of the angle of incidence θ. From the angular relationship presented previously, the following functions are obtained.

$$f_1 = \frac{\cos(a_j + b_j \theta)}{\cos\theta} \left|\underset{\sim}{\Omega} \times \underset{\sim}{R}\right|^{-2} \tag{5}$$

$$f_2 = \sin(a_j + b_j \theta).$$

Equations (1) and (3) can then be integrated to find the torques and drag forces on the satellite in terms of the parameters σ_j, α_j, a_j, b_j and ρ.

6.2.3.7. Aerodynamic Properties of Spinning Bodies

The results of the aerodynamic properties study of spinning bodies are presented using the following notations and definitions.

Drag = force in the direction of the velocity $\underset{\sim}{U}_\infty$.

$(\text{Lift})_Z$ = force normal to $\underset{\sim}{U}_\infty$ and in the plane formed by $\underset{\sim}{U}_\infty$ and the spin vector, $\underset{\sim}{\Omega}$.

$(\text{Lift})_X$ = lateral force perpendicular to $\underset{\sim}{U}_\infty$, $\underset{\sim}{\Omega}$, and $(\text{lift})_Z$.

T_s = slow-down torque in the direction of $\underset{\sim}{\Omega}$.

T_j = component of precession torque perpendicular to $\underset{\sim}{\Omega}$ and in the plane formed by $\underset{\sim}{U}_\infty$ and $\underset{\sim}{\Omega}$.

T_i = component of precession torque perpendicular to $\underset{\sim}{U}_\infty$, $\underset{\sim}{\Omega}$, and T_j.

θ_s = $\frac{1}{2}\pi$ minus the angle between $\underset{\sim}{U}_\infty$ and $\underset{\sim}{\Omega}$.

$k_i = r_i\Omega/U_\infty$, where r_i is the maximum radius about the spin axis of the i^{th} body.

$I_{\ell k}$ = integral functions of the gas surface interaction parameters (a_j, b_j, σ_j, and α_j), k_i, θ_s and the variables of integration.

With the above definitions, the results of the aerodynamic properties study are presented as follows.

Spinning Disk at Angle of Attack

Let $D_D = \rho U_\infty^2 \sin\theta_s r_d^2$, where ρ is the gas density and r_d the radius of the disk.

$$\text{Drag} = D_D[\pi - I_{11}] \qquad T_s = -D_D r_d[k_d(\pi/2) + I_{14}]$$

$$(\text{Lift})_Z = D_D[I_{12}] \qquad T_j = D_D \sin\theta_s r_d[I_{15}]$$

$$(\text{Lift})_X = D_D[I_{13}] \qquad T_i = -D_D \sin\theta_s r_d[I_{16}]$$

Spinning Cylinder at Angle of Attack

Let

$$D_{cy} = \rho r_{cy} L U_\infty^2 \cos\theta_s$$

where r_{cy} is the radius of the cylinder and L the length.

$$\text{Drag} = D_{cy}[2 + I_{21}] \qquad T_s = D_{cy} r_{cy}[2k_{c1} + I_{24}]$$

$$(\text{Lift})_Z = D_{cy}[I_{22}] \qquad T_j = D_{cy} r_{cy} \sin\theta_s[I_{25}]$$

$$(\text{Lift})_X = -D_{cy}[\tfrac{1}{2}k_{cy}\pi] \qquad T_i = D_{cy} r_{cy} \sin\theta_s[\tfrac{1}{2}\pi - I_{26}]$$

Spinning Cone at Angle of Attack

The results for the cone are divided into two cases dependent upon the angle of attack and the cone half angle, δ.

Case 1: When $(\tan\theta_s)(\tan\delta) \geq 1$, the entire surface of the cone is exposed to the flow. The following results are applicable for this case.

Let

$$D_c = \rho U_\infty^2 H_c r_c / \cos\delta$$

where H_c is the height and r_c the base radius of the cone.

$$\text{Drag} = D_c [\pi \sin\theta_s \sin\delta + I_{31}]$$

$$(\text{Lift})_Z = D_c [I_{32}]$$

$$(\text{Lift})_X = -D_c [k_c (\pi/3) \cos\theta_s \cos\delta + I_{33}]$$

$$T_s = D_c [r_c k_c (\pi/2) \sin\theta_s \sin\delta + I_{34}]$$

$$T_j = -D_c [H_c k_c (\pi/12) \cos\theta_s \cos\delta + I_{35}]$$

$$T_i = D_c [I_{36}]$$

Case 2: When $-1 \leq (\tan\theta_s)(\tan\delta) \leq 1$, part of the cone surface is shielded from the flow. The following results are applicable for this case where,

$$\beta = \sin^{-1}(\tan\theta_s \tan\delta),$$

and where D_c, H_c, and r_c are defined as in case 1.

$$\text{Drag} = D_c [\cos\theta_s \cos\delta \cos\beta + \tfrac{1}{2} \sin\theta_s \sin\delta (\pi + 2\beta) + I_{41}]$$

$$(\text{Lift})_Z = D_c [I_{42}]$$

$$(\text{Lift})_X = D_c [(2/3) k_c \sin\theta_s \sin\delta \cos\beta - (1/6) k_c \cos\theta_s \cos\delta (\pi + 2\beta - 2\text{in}2\beta + I_{43}]$$

$$T_s = -D_c [(1/4) r_c k_c (\sin\theta_s \sin\delta (\pi + 2\beta) + 2\cos\theta_s \cos\delta \cos\beta) - I_{44}]$$

$$T_j = -D_c [(1/12) H_c (\tfrac{1}{2} \cos\theta_s \cos\delta (\pi + 2\beta - \sin 2\beta) + 2\sin\theta_s \sin\delta \cos\beta + I_{45}]$$

$$T_i = D [(2/3) r_c \sin^2\theta_s \sin\delta \cos\beta - (1/6) r_c \sin\theta_s \cos\theta_s \cos\delta \sin^2\beta$$
$$- (1/3) H_c \cos^2\theta_s \cos\delta \cos\beta + I_{46}]$$

The terms $I_{\ell k}$ are surface integrals of the force produced by the reflection part of the gas surface interaction. These terms are currently being evaluated by the numerical technique described in previous reports.

For special cases of the reflection models, the $I_{\ell k}$ terms can be evaluated analytically. These cases are

1. Diffuse reflection

2. Specular reflection with or without energy accommodation.

With these two cases, the Maxwell model can be constructed and also a restricted case of the Nocilla model and the Schamberg model. Since analytic solutions for the aerodynamic properties can be obtained by using the above special cases, these results will be used first in the calculations of the properties of a composite satellite.

6.2.3.8. Numerical Techniques

The procedure of analysis being followed is to first evaluate the aerodynamic properties of selected satellite shapes as a function of angle of attack. The results obtained from this program will then be used as the necessary input data for a computer program which evaluates the spin slow-down rate, precession rate, and drag effects as a function of the orbital parameters.

Since the computation of the aerodynamic properties of a given body shape requires the evaluation of numerous double integrals, some of which have no closed-form solution, a two-dimensional quadrature formula is being used to evaluate the integrals. The following 9-point approximation is currently being used for each square subdivision of the domain of integration.

$$\int_D\int f(x,y)dxdy \approx 4h^2 \sum_{i=i}^{9} \omega_i f(x_i,y_i).$$

The formula, abscissas (x_i,y_i) and weights, ω_i, are taken from Ref. [12]. The 9-point quadrature formula used is exact for mixed polynomials of degree 5 or less with remainder of order h^6. The accuracy of the formula was found to be adequate for the integration of the type of functions which arise in this study.

The equations for the disk, cone, truncated cone, and cylinder are being evaluated systematically using the numerical method shown above. However, to integrate the drag and torque equations for the sphere and spherical

segment, a modification of the present numerical quadrature formula must be made to allow integration over curved domains. Numerical integration formulas such as those given by Tyler [13] appear to be more applicable to problems of curved boundaries than the present formula. Tyler's formula for rectangular domains provides the degree of flexibility in application needed for this work. The error due to the neglecting of small areas near the edge of the domain is currently being studied and methods of improving the accuracy of the calculation are to be developed.

G.R. Karr

6.2.4. Effects of Degassing and Adsorption

The effects of adsorbed gases on the aerodynamics of the proposed satellite are being studied. The angular distribution of atoms reflected from a metal surface has been shown to depend on the quantity of gas adsorbed on the surface [14]. Since adsorption and desorption are time dependent until a steady state condition is achieved, the possibility of measuring the desorption or adsorption rate is being studied.

There are two principle types of adsorption: physical adsorption in which van der Waals attraction between the gas and solid surface is the primary mechanism, and chemical adsorption where the attraction comes from unsatisfied valency requirements of the surface atoms. The characteristic energy of physical adsorption is on the order of one to two times the latent heat of vaporization of the gas, while for chemisorption the energies are about the same as those that would be associated with the formation of a chemical compound. Physical adsorption is reversible while chemisorption is usually irreversible.

The composition of the gases residing on the satellite surface at the beginning of the experiment includes effects due to the prelaunch history of the satellite material as well as what changes occur during the launch into the space environment. For metals some typically adsorbed gases are N_2, H_2, CO_2, and O_2 on iron and tungsten, and CO and O_2 on aluminum. Glasses will absorb water vapor with CO_2 and N_2 next in abundance [15].

At a typical altitude for this experiment, 300 km, the atmosphere is composed of atomic oxygen, diatomic nitrogen and oxygen, in decreasing

order of abundance. Some of these molecules can be expected to adhere to the surface while others may eject existing adsorbed atoms. Also, free electrons, ions, cosmic rays, and electromagnetic radiation interact with the surface of the satellite to alter the rate of desorption. These effects are being studied to find if they are significant in affecting the satellite aerodynamics.

6.2.4.1. Degassing Effects

It is found that there are combinations of gases and metals which have degassing rates that are compatible with the interval of time necessary to observe the satellite and that this rate is strongly dependent on surface temperature.

This analysis is limited to monolayer adsorption, since secondary gaseous layers, having small values of latent heat of vaporization, are rapidly evaporated even at very low temperatures. All other particle interactions with the gaseous layer (e.g., collisions with neutral molecules, ions, electrons, photons, and cosmic rays) will be neglected. Except for the adsorbed gas, the surface is assumed to contain no other contaminants and to be homogeneous throughout. This requirement implies that E, the activation energy of desorption, does not vary over the surface. Finally we will neglect interactions between adsorbed molecules. Thus E will be independent of the amount of surface coverage.

The fractional surface coverage, θ, for the constant residence time approximation is given by $\theta = e^{-t/\tau_r}$, where $\tau_r = \tau_o e^{E/RT}$ is the residence time or the average time an adsorbed molecule remains on the surface, τ_o is the period of the molecular oscillations normal to the surface, approximately 10^{-13} sec, E is the activation energy of desorption, R is the molar gas constant, and T is the absolute temperature. For constant surface temperature, τ_r is a constant because of the assumption of constant E.

The range of surface temperature of the proposed satellite can be controlled by choosing construction materials having the proper value of α/ϵ, the ratio of absorptivity to emissivity. For example, a sphere in Earth orbit with $\alpha/\epsilon = 0.1$ would reach an equilibrium temperature of 150K. A flat plate, with its surface normal pointing towards the sun and with a value of $\alpha/\epsilon = 9$, will reach 600K. The equilibrium temperatures reported are those due to exposure to solar radiation and do not take satellite eclipses into account.

In order to have a measurable amount of degassing, require the

fractional surface coverage to be down to 1/e of the initial coverage in the time t. Furthermore, the observing time interval will be limited to be between 1 and 10 weeks. Thus, the satellite must conveniently degass to 1/e of its initial surface coverage in a time greater than 1 week but less than 10 weeks, i.e., 1 week $< \tau_r <$ 10 weeks. Since τ_r and T are both bounded (150K $<$ T $<$ 600K), the range of E values necessary to give the proper desorption rate can be determined. Values of E between 0.6 eV and 2.4 eV per molecule are found to give a measurable degassing rate. Table 6.1 lists some gas-metal systems with activation energies of desorption in this range. Of these, H_2O on metal surfaces may be the most promising because it is prevalent on surfaces exposed to sea-level atmosphere.

The preceding discussion treats the idealized case of adsorption using a constant τ_r. However, solid surfaces are not homogeneous, but contain cracks, grain boundaries, and micropores. These imperfections give rise to a distribution of desorption energies. Furthermore, if interactions between adsorbed molecules are considered, E is found to depend on the amount of surface already covered. Thus, for a perfectly clean surface, the first adsorbed molecules encounter maximum E while the last few molecules to form the monolayer encounter minimum E. Also, the surface temperature cannot be expected to remain constant since the satellite's motion will continually expose different surfaces to the sun. All these factors combine to give a varying τ_r.

Table 6.1. Activation Energy of Desorption for Some Gas-Metal Systems

Gas-Metal System	E (eV per molecule)
Maximum heat of physical Adsorption of H_2O [16]	0.61
H_2O on Metals [17]	0.95 - 1.04
H_2 on W	2.0
N_2 on Fe	1.74
CO on Ni	1.52
CO on Fe	1.39

6.2.4.2 Gas Adsorption Effects

The adsorption of gases on the proposed satellite is discussed with respect to the possibility of measuring their rate of adsorption. The desorption of a satellite surface was discussed in section 6.2.4.1.

The fractional coverage θ at the time t of an initially clean surface using the Langmuir expression for the equilibrium concentration is given by

$$\theta = [\alpha_o \upsilon/(\alpha_o \upsilon + N_m/\tau_r)]\{1-\exp[-(\alpha_o \upsilon/N_m + 1/\tau_r)t]\},$$

where α_o = condensation coefficient

υ = the collision rate per unit area of molecules with the surface

N_m = the number of molecules needed to form a monolayer

τ_r = average time a molecule remains on the surface $\approx \tau_o \exp(E/RT)$

In order for the adsorption to be measurable, two requirements must be satisfied: (1) the fractional coverage must be large enough to alter the aerodynamic properties of the satellite, and (2) the time for the fractional layer to form must be on the same order as the time needed to make suitable measurements of the forces and torques.

The first requirement can be fulfilled by placing a lower limit on the fractional coverage, such as

$$\theta > 0.01 \text{ as } t \to \infty$$

The second requirement is satisfied if the time at which θ reaches 63% of its final value is limited by:

$$1 \text{ week} < \left(\frac{\alpha_o \upsilon}{N_m} + \frac{1}{\tau_r}\right)^{-1} < 10 \text{ weeks}$$

Letting $\alpha_o = \frac{1}{2}$, $N_m \approx 5\times10^{18}\ m^{-2}$, and υ at 300 km $\approx 8\times10^{18}\ m^{-2}s^{-1}$, there obtains

$$\frac{\alpha_o \upsilon}{N_m} = 0.8s^{-1}$$

With this calculation, the time is found to be too short to be measured for any value of τ_r. In addition, there appears to be no reasonable re-adjustment of parameters which would allow such a measurement to be made.

A. Cohen

6.3. Infrared Fourier Interferometer[†]

Absorption spectra to determine the nature of molecular vibrations of cyclododecasulfur, S_{12} and to study the sulfur-sulfur band S-S, were measured with the IR interferometer in cooperation with the Chemistry and Chemical Engineering Department.

The form S_{12} was first prepared by Prof. M. Schmidt and Dr. E. Wilhem in 1966 [18], and its crystal structure has been studied by Dr. A. Kutoglu and Prof. E. Hellner by X-ray diffraction [19]. It is a 12-membered ring with the sulfur atoms lying in 3 planes. It forms small, needle-type crystals and is very stable in air at room temperatures. The laboratory sample was obtained from Prof. M. Schmidt.

A solid sample was first prepared by pressing the S_{12} crystals in an evacuated die into a disk approximately 13 mm in diameter and 0.5 mm thick. Absorption spectra were obtained in the 2 to 40 $\times 10^3$ m^{-1} range at a resolution of $200m^{-1}$. Absorption bands were found at 16.7, 17.3, 25.4, and 27.1 $\times 10^3$ $m^{-1} \pm 0.4 \times 10^3$ m^{-1}.

Absorption spectra of the same sample using a Beckman IR-11 grating spectrometer with a resolution of $200m^{-1}$ were obtained, giving absorption bands at 16.6, 17.3, 25.2, 26.8 and 47.0 $\times 10^3$ m^{-1}.

In order to help the assignment of fundamental frequencies, polarized absorption spectra using both the interferometer and spectrometer were used. The molecular principal axis (Z-axis) was found to be along the crystal axis by X-ray rotational photographs. A thin film of fluorolube oil was spread on a polyethylene sheet and the S_{12} crystals were laid on the sheet with their crystal axes all along one direction. A disk sample of about 13 mm in diameter was prepared this way. A polarizer made of two polished parallel silicon plates [20] was used in these absorption measurements. The polarizer was placed in front of the S_{12} sample and absorption spectra of parallel and perpendicular polarizations were obtained with the results shown in Table 6.2.

Table 6.2. Polarization Measurements

Frequency - 10^3 m^{-1}	16.6	17.3	25.2	26.8	47.0
Interferometer			$\perp$	$\parallel$	$\parallel$ and $\perp$
Spectrometer	$\perp$	$\perp$	$\perp$	$\parallel$	

[†]Supported in part by the Joint-Services Electronics Program under contract number DAAB-07-67-C-0199.

The two techniques were complementary, with the interferometer measuring spectra in the far infrared region and the spectrometer in the 20 micron region. The spectrometer could not be readily evacuated so that rotational spectra of water and air were present. The interferometer was evacuated and produced spectra with high signal-to-noise ratios.

Fang-maw Lin
D. Skaperdas

6.4. References

1. Progress Report for March through August, 1968, Coordinated Science Laboratory, University of Illinois, pp. 133-134, (September 30, 1968).

2. James L. Myers, "Optimization of a Gyroscopic Satellite General-Relativity Experiment," Technical Report R-396, Coordinated Science Laboratory, University of Illinois, (September 1968).

3. D.H. Cooper, G.R. Karr, J.L. Myers, and D. Skaperdas, "A Proposed Test of the Einstein Theory of Gravitation by Means of an Unshielded Orbiting Gyro Using Passive Telemetry," Technical Report R-378 (Final Report under NASA Research Grant NsG-443), Coordinated Science Laboratory, University of Illinois (May 1968).

4. R.H. Dicke, "Solar Oblateness and General Relativity," Phys. Rev. Letters 18, 313 (Feb. 27, 1967).

5. L.I. Schiff, "Motion of a Gyroscope According to Einstein's Theory of Gravitation," Proc. Natl. Acad. Sci. U.S. 46, 871 (1960).

6. C. Brans and R.H. Dicke, "Mach's Principle and a Relativistic Theory of Gravitation," Phys. Rev. 124, 925 (1961).

7. P. Rastall, "An Improved Theory of Gravitation," Can. J. Phys. 46, 2155 (1968).

8. P. Rastall, "Relativistic Precession of a Spinning Satellite," Can. J. Phys. 44, 3109 (1966).

9. R. Schamberg, "A New Analytic Representation of Surface Interaction with Hypothermal Free Molecule Flow with Application to Neutral-Particle Drag Estimates of Satellites," Rand Research Memorandum RM-2313, January 8, 1959, Rand Corporation, Calif.

10. Silvio Nocilla, "The Surface Re-Emission Low in Free Molecule Flow," Rarefied Gas Dynamics, Third Symposium, ed. by J.A. Laurmann, 1963, Vol. 1, pp. 327-346.

11. G.S. Reiter and K. Moe, "Surface-Particle-Interaction Measurements Using Paddlewheel Satellites," Rarefied Gas Dynamics, Sixth Symposium, 1968, in press.

12. M. Abramowitz and I. Stegun, Handbook of Mathematical Functions, pp. 892-893, Dover Publications, Inc., New York (1965)

13. G.W. Tyler, "Numerical Integration of Functions of Several Variables," Can. Jour. of Math., 3, 393 (1953).

14. D.R. O'Keefe and J.B. French, "High Energy Scattering of Inert Gases from Well Characterized Surfaces. I Experimental," Rarefied Gas Dynamics, Sixth Symposium, 1968, in press.

15. S. Dushman and J.M. Lafferty, Scientific Foundations of Vacuum Technique, John Wiley and Sons, Inc., New York, 1962.

16. B.M.W. Trapnell, Chemisorption, Academic Press, Inc., New York, 1955.

17. D.J. Santeler, D.H. Holkebaer, D.W. Jones, F. Pagano, Vacuum Technology and Space Simulation, NASA SP-105, Washington, D.C., 1966.

18. M. Schmidt and E. Wilhelm, Angew. Chem. Internat. Edit 5, 964 (1966).

19. A. Kutoglu and E. Hellner, Angew. Chem. Internat. Edit 5, 965 (1966).

20. C.J. Johnson, Master Thesis, Electrical Engineering Dept., University of Illinois, 1968.

7. SEMICONDUCTOR PHYSICS[†]

W.D. Compton
B. Arora
D.E. Cullen
C.E. Jones
E.S. Johnson
H.V. Krone
H. Peisl
L. Schein

7.1. Introduction

Defects, either impurities or intrinsic lattice defects, often are critical in determining many of the properties of semiconductors. Irradiation with high-energy particles is a convenient way of introducing simple defects into solids. These elemental defects, vacancies and interstitials, diffuse through the lattice and interact with impurities and with each other to form a variety of more complex defects. Since these defects can serve as efficient recombination centers for electrons and holes, the luminescence resulting from this recombination is a useful tool for studying the position of the energy level in the gap that results from the defect, the electron-phonon interaction that occurs with charge trapped at the defect, and the microscopic symmetry of the defects. These techniques are being utilized in the study of irradiated silicon.

The II-VI semiconductors can be excited into laser emission by strong optical excitation. Defects affect both the wavelength of the emission of these materials and the efficiency of the laser action. High-resolution spectra of the laser luminescence is being used as a tool for studying the influence of the level of excitation, the temperature, and the defect concentration upon the laser action.

P-i-n diodes have been shown to be efficient infrared detectors. Recombination centers arising from impurities and from radiation-induced defects have been shown to influence the oscillatory behavior of these defects.

Tunneling between a metal and a semiconductor is an excellent tool for examining many of the properties of the degenerate semiconductor, including electron-phonon interactions, density of states, and surface-barrier effects. Results have been obtained on p-type silicon where the acceptor concentration has been significantly varied.

[†]Supported principally by the National Aeronautics and Space Administration through Jet Propulsion Laboratory Contract No. 952383; and in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy and U.S. Air Force) under contract DAAB-07-67-C-0199 and the Physical Electronics Affiliates Program, College of Engineering, University of Illinois.

7.2. Recombination Luminescence in Irradiated Silicon

7.2.1. High-Resolution Spectra

Figure 7.1 reproduces data that were taken by Robert Spry of the spectrum of the recombination luminescence of silicon exposed to Co^{60} radiation equal to that which would produce 2.6×10^4 C/kg of ionization in air (10^8 rontgen) [1]. As indicated on the figure, a resolution between 0.010 and 0.005 eV was used for this measurement. On the basis of the energy of the peaks from the principal zero-phonon lines, A or C, the lines labeled B, D and E were tentatively assigned to transitions that result from phonon-assisted transitions with TA and TO phonons involved.

The results of the higher-resolution measurements are shown in partial form in Fig. 7.2. Lines B, D and E are presented with a spectral resolution of between 0.001 and 0.0005 eV, a factor of about ten better than was used in Fig. 7.1. The importance of this measurement is the observation that these lines have the same half width as the zero-phonon lines A and C. Thus, it is likely that these are also zero-phonon lines and that they do not arise from phonon-assisted transitions.

The earlier data were taken with the sample held to a copper sample holder with vacuum grease. As will be shown below, this can introduce substantial strain into the crystal, thereby tending to broaden and shift the peak position of some of the sharp lines. Lines A and C, when measured without stress, are located at 0.975 and 0.795 eV, respectively. The difference in the position and width of line B between the previous and present measurements may be due to the same effects. Additional data are being taken to aid in the determination of the source of these lines.

7.2.2. Uniaxial Stress

The difference in thermal contraction between silicon and metal is used to produce uniaxial stress in silicon at low temperatures. Silicon has $\Delta\ell/\ell = 2.6 \times 10^{-3}$ between 300 K and 0 K, brass has $\Delta\ell/\ell = 4 \times 10^{-3}$, and 304 stainless steel has $\Delta\ell/\ell = 3 \times 10^{-3}$. The silicon samples are orientated by x-ray diffraction to within about $\pm 2^{\circ}$ with respect to the crystal axis. They are cut into long rectangular pieces 18.3X1.5X3mm, with the length corresponding to the <100>, <110>, or the <111> crystal directions. The ends of the samples are coated with metal and Thermit-Corporation number 323 red lacquer. The

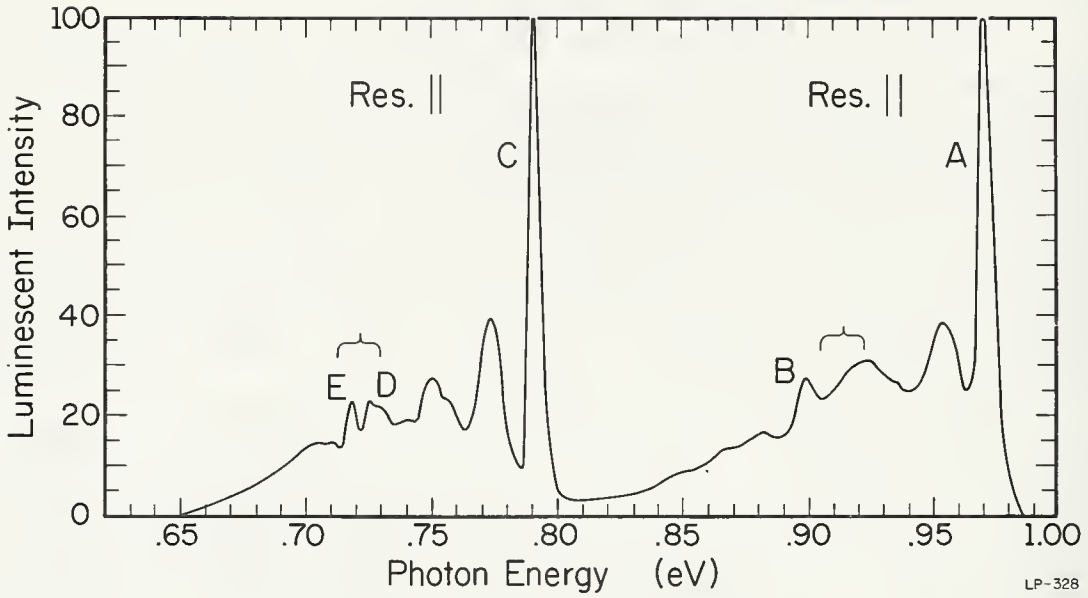


Fig. 7.1. Recombination luminescence in Co^{60}-irradiated 1.0Ωm N-type silicon.

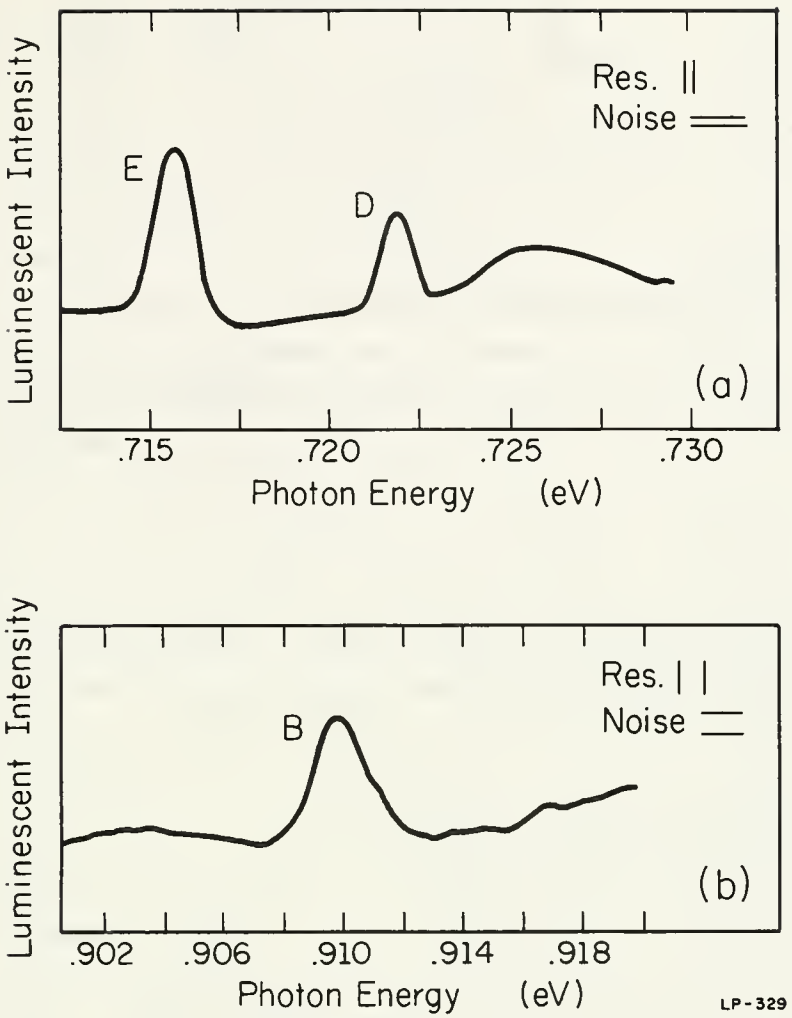


Fig. 7.2. High resolution spectra of lines labeled B, D and E of Fig. 7.1.

exposed silicon is etched in CP-8. The red lacquer is peeled off, and the sample is mounted into a piece of brass or stainless stell which has a section of 18.3X1.5X3.2mm milled out. A part over the center etched section of the crystal allows the luminescence to be collected and measured. The metal holder is attached to the tail piece of a variable-temperature, optical, helium dewar. On cooling to near liquid-helium temperature, the metal contracts more than the silicon applying a stress along the length of the sample.

The effects of stress upon the band-to-band luminescence in silicon are shown in Fig. 7.3. Curve a is the band-to-band luminescence with no applied stress. Curves b to e present the effects of stress and temperature. The stress is essentially compressional in the <100> plane. The light is incident along <001>. Curves b to e are taken at 4.6, 10, 19 and 32K respectively. Note that the stress results in a splitting of the line and a shifting of the luminescence to lower energy. This result is completely consistent with the splitting and shifting of the silicon conduction band relative to the valence band when a uniaxial stress is applied. A comparison of these results with those of Bulshve [2] indicates that the splitting arises from a stress of approximately 6×10^7 N/m^2.

The data for the zero-phonon line at 0.789 eV, shows the line splitting into 2 components for stress in the <100> direction, 2 components for the <110> direction and one component for the <111> direction. This would imply that the defect has tetragonal <100> symmetry [3]. The data for the peak at 0.969 eV is less reliable for the monochromater grating passes only vertically polarized light in this region. Peaks strongly polarized perpendicular to the stress direction will be strongly suppressed. The splittings appear to give one component for stress in the <100> direction, two components for the stress in the <110> direction, and one, or possibly two, components for the stress on the <111> direction. This would correspond to a defect with a trigonal <111> orientational symmetry. Both of these assignments must be viewed as tentative, since they are at present based on only a small number of experiments.

7.2.3. Effect of Lithium Impurity

The material used in these experiments is given in Table 7.1.

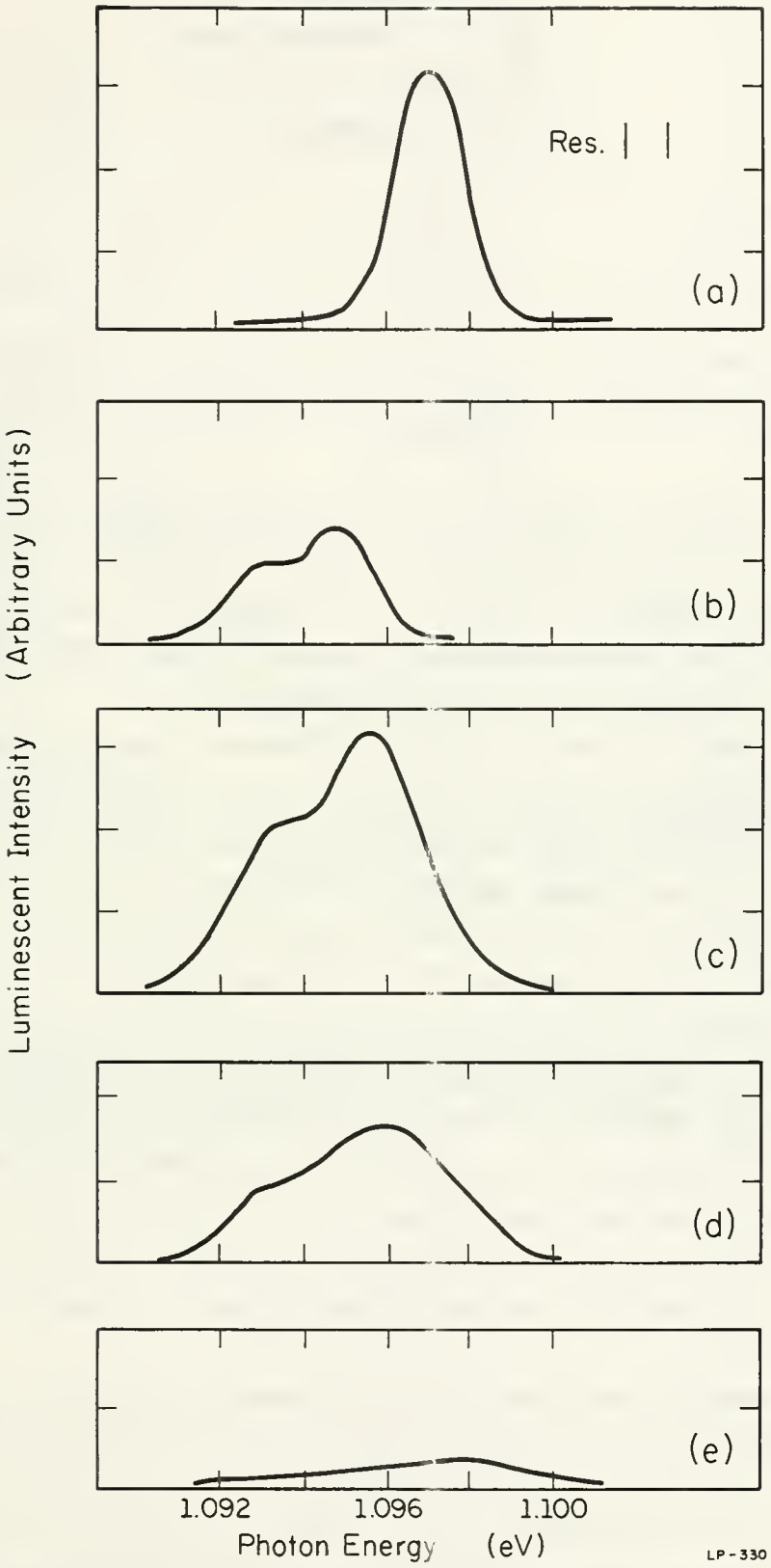


Fig. 7.3. Effect of compressional stress in the ⟨100⟩ plane upon the band-to-band luminescence in silicon, 0.90Ωm P-type pulled. (a) No stress; (b)-(e) Effect of compressional stress in ⟨100⟩ plane; b) 4.6K, c) 10K, d) 19K, e) 32K.

Table 7.1 Material Used

	Resistivity Prior to Li Diffusion	Lithium Content
N-type Float Zone	0.70 Ωm	$5.5\times10^{23}/m^3$
N-type Pulled	1.00 Ωm	$1.5\times10^{24}/m^3$
P-type Float Zone	0.65 Ωm	$1.2\times10^{24}/m^3$
P-type Pulled	0.45 Ωm	$1.1\times10^{24}/m^3$

Samples for each of these have been irradiated with 10^{20}, 10^{21} and 10^{22} e/m^2 having an energy of 3 MeV. Luminescent spectra have been taken after each irradiation with the lower-resolution system described above. Note that the representative spectra presented below were taken before the improvement was made in the detector assembly of this system.

Figure 7.4 shows the effect of lithium upon the recombination luminescence of n-type float zone material after irradiation with essentially 4×10^{22} e/m^2. Curve A, taken on material without lithium, presents the usual spectrum with a zero-phonon line at 0.97 eV, and the family of peaks at lower energy and the line at 0.897 eV that we know from the above are also zero-phonon lines. The introduction of the lithium results in a tremendous broadening of this spectrum, and the introduction of two new luminescent lines into the spectrum at 1.027 and 1.046 eV as shown in Curve B. The zero phonon line at 0.97 eV is still evident on the large background. The structure at 1.027 and 1.046 is always seen in n-type float zone, p-type float zone and n-type pulled material that contains lithium at the high irradiation levels. It has never been seen in p-type pulled and is never seen in any material free of lithium. In p-type float zone and n-type pulled it is seen after irradiation with 10^{21} e/m^2.

Although these results are preliminary, the lithium appears to introduce a recombination level that is much closer to the band edge than is found in the lithium-free material. Thus, a luminescence has been observed that appears to be determined by the lithium. Its properties, the relationship of the defect responsible for it, and the defects responsible for the other luminescent emission are yet to be determined.

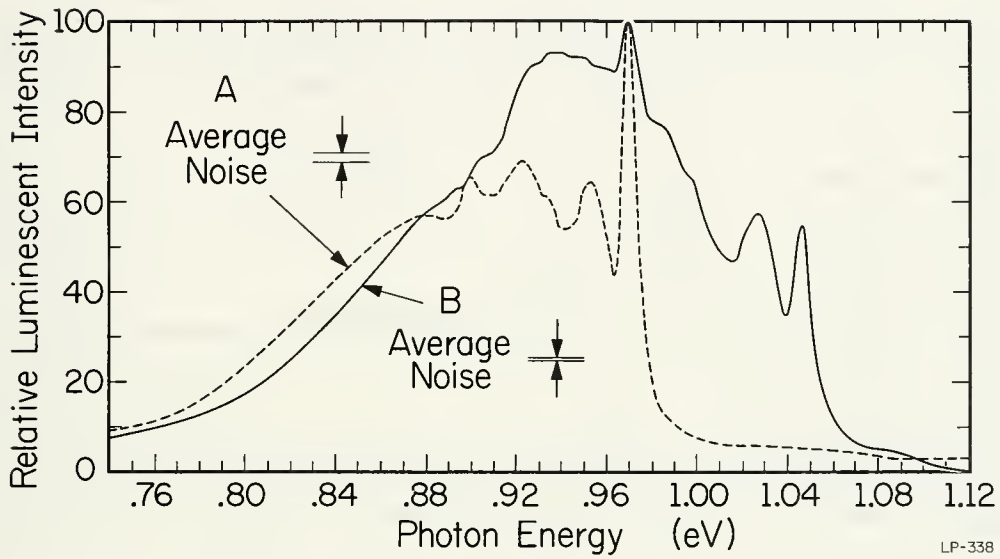


Fig. 7.4. Recombination luminescence of irradiated silicon. (A) N-type float zone 0.70Ωm, irradiated with 4.2×10^{24} e/m^2; (B) N-type float zone 0.70Ωm with 5×10^{23} /m^3 of lithium. Irradiated with 3.5×10^{24} e/m^2.

Absolute measurements of luminescent intensity as a function of irradiation level are very difficult to make, as a result of the etching techniques that must be used. A threshold appeared to exist for the irradiation needed to produce luminescence. The intensity of the luminescence irradiations below 10^{20} e/m^2, if present at all, was barely above the noise level. It is hoped that the increased sensitivity of the system will allow a better measure of this threshold effect.

A somewhat different effect of the lithium is seen in Fig. 7.5 for an n-type pulled sample with lithium after irradiation with 10^{21} e/m^2 of 2 MeV. Note that the instrumental resolution is substantially poorer in Fig. 7.5 than in Fig. 7.4. Of particular interest in this case is the luminescence at low energy that appears to show no structure, at least within this noise level and resolution. Additional measurements are certainly needed to establish the origin of this spectrum.

Although luminescent spectra have been measured on electron-irradiated samples of all of the materials prior to lithium doping that are listed in the table above, only curve A of Fig. 7.4 is presented for the non-lithium-doped material. Most of the detailed features of the data, as presented by Spry and Compton [1] for neutron and gamma-ray bombardment, have been confirmed for electron irradiations. Several observations have been made, however, that will need to be examined more closely with the more sensitive system. For each material type, the intensity of the luminescence appears to be significantly lower after irradiations with 10^{22} e/m^2 that it was after irradiation with 10^{21} e/m^2. In n-type pulled, significant luminescence resulting from recombination via the defects was seen after irradiation with 10^{18} e/m^2, 3 MeV. Thus, the sensitivity of formation of the defects giving rise to the 0.97 eV and 0.79 eV family of curves is substantially greater than for the same material after lithium diffusion. Finally, a sharp line at about 10nm higher energy than the 0.79-eV zero-phonon line has been seen in some samples. This line has also been seen irregularly in gamma-ray and neutron irradiated samples. It is not clear at this time what the relevant parameter is for the observation of this line. Similarly, a luminescence peaking at about 0.5 eV has also been observed in some samples of the irradiated n-type pulled material.

The above data have been presented as examples of the survey data

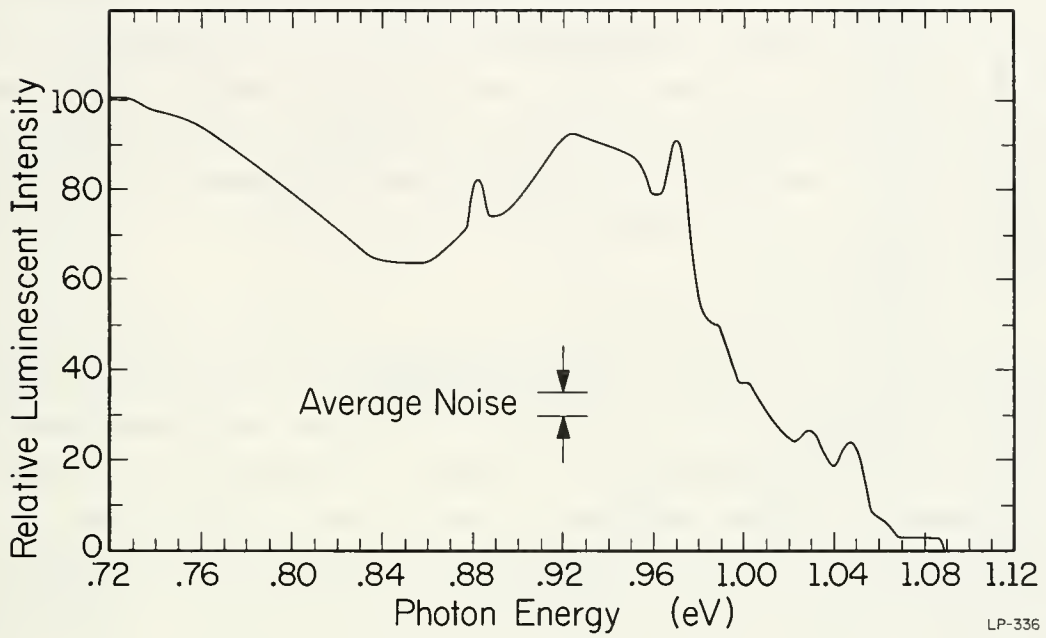


Fig. 7.5. Recombination luminescence on irradiated N-type pulled silicon containing 1.5×10^{24} /m^3 of lithium. Irradiated with 10^{23} e/m^2.

that have been taken. The availability of the more sensitive detection system will require that much of these data be taken again so that more definitive answers can be obtained for many of the questions that were left unanswered in the above discussion.

7.2.4. Effect of Thermal Annealing

A survey of the effect of high-temperature anneals upon the recombination spectra of the irradiated material has not been completed. Annealing data have been taken on n-type float zone following irradiation with both 10^{20} and 10^{21} e/m^2 and n-type pulled following irradiation with 10^{20} e/m^2. Although no systematics of the effects of high-temperature annealing can be presented at this time, an example of the effects of such an anneal will be given for n-type pulled material.

The luminescent intensity of an n-type pulled sample was monitored for about seven weeks after irradiation. Both 0.97 eV and 0.79 eV peaks were visible just after irradiation with the 0.79 eV peak being about a factor of two lower in intensity than the 0.97 eV peak. It was found that upon standing at room temperature, that the ratio of the 0.79 eV to the 0.97 eV peak increased to about unity and then gradually decreased to less than 0.1. The spectrum was measured at this time and is shown in Fig. 7.6. The sample was then annealed at 300°C for 14 hours and the spectrum of Fig. 7.7 obtained. A tremendous increase in the 0.79 eV peak is seen. The insert, taken at higher resolution shows the line at 10 nm higher energy than the 0.79-eV line. Notice also that the 0.97-eV zero-phonon line no longer seen and that a new family of lines has appeared at slightly lower energy. It is not clear whether this new luminescence results from a new defeat created by the anneal, or whether this series has simply become evident because of the absence of the stronger 0.97-eV family.

One n-type float-zone crystal also has been annealed. The only luminescence seen before anneal, the 0.97-eV zero-phonon family, is destroyed and a weak structureless band was seen, stretching from the monochromator's low-energy limit of 0.70 eV. It is expected that the systematic investigation of the affects of annealing will yield additional information on the relationship of the defects that give rise to the two luminescent patterns at 0.79 and 0.97 eV.

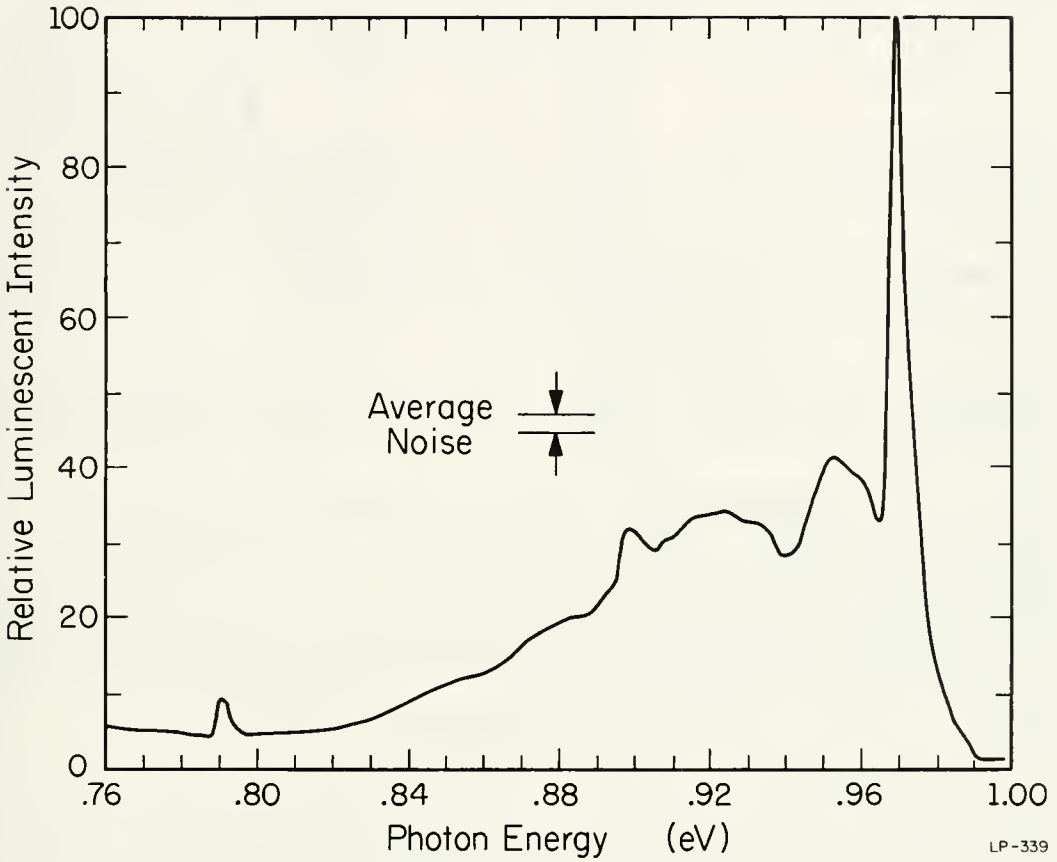


Fig. 7.6. Recombination luminescence of 1.0Ωm N-type pulled silicon after irradiation with 10^{22} e/m^2 and room temperature anneal for seven weeks.

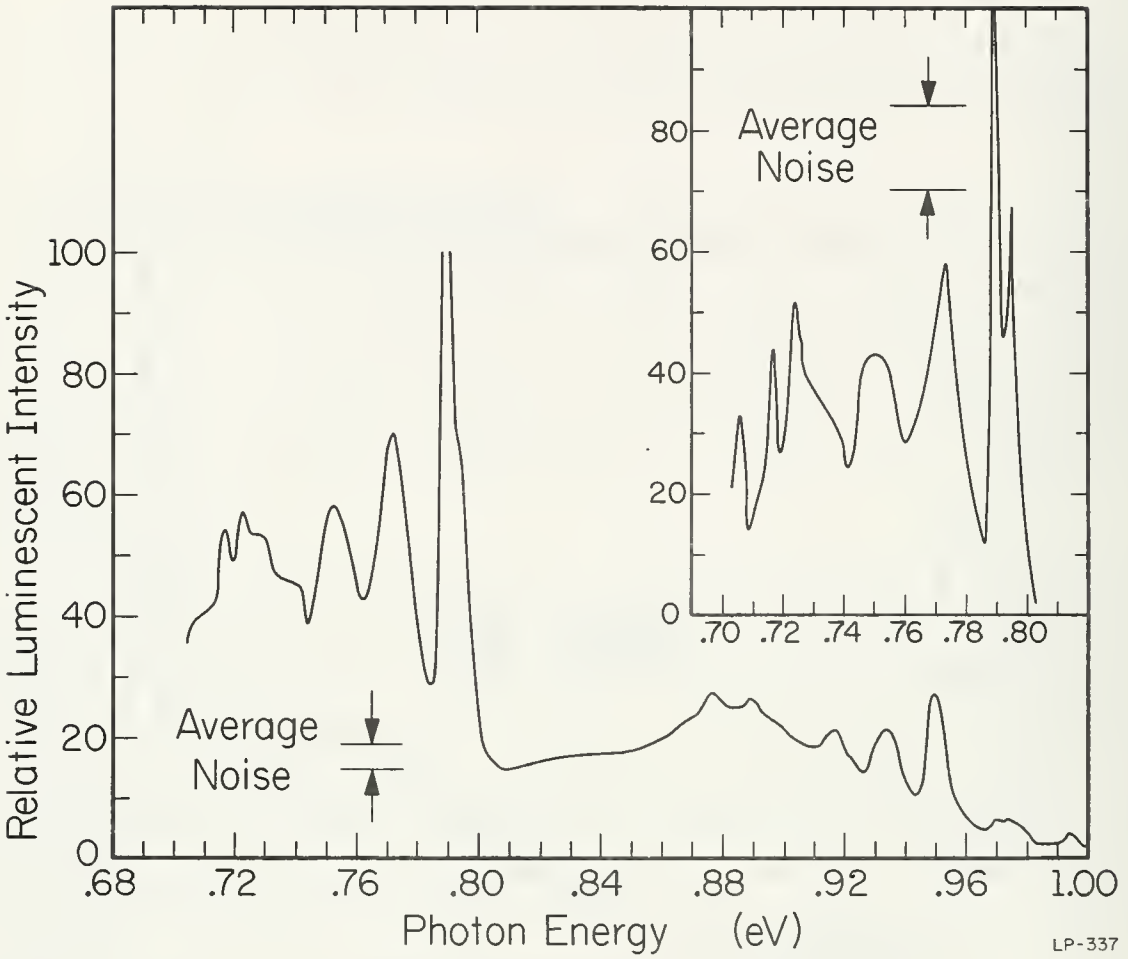


Fig. 7.7. Recombination luminescent spectrum of sample used in Fig. 7.7 after an anneal at 300°C for 14 hours.

7.3. Radiation Effects in Semiconducting Laser Materials[†]

7.3.1. Instrumentation

A Jarrell-Ash, Model 78-400, 1.8-meter grating spectrometer was acquired by the Laboratory for use with this project. This instrument, with a resolving power of 500,000, is ideal for the types of measurements required for this study The spectrometer and the necessary detector and recording circuitry have been assembled and calibrated.

7.3.2. Materials

High quality platelets of Cd(S,Se) that can be excited into laser action have been grown by Argon-vapor transport. The source material was obtained from Eagle-Pitcher (CdSe) and from Sylvania (CdS). Laser action has been seen at 77K and at 4K with a groupd of CdSe platelets when excited by a pulsed Ga(AsP) laser diode.

The Ga(AsP) laser diodes were produced from Ga(AsP) material using Zn as dopant.

7.3.3. Effects of Radiation Upon Ga(AsP) Laser Diode

A number of high-resolution measurements were made of the emission of a commercial Ga(AsP) laser diode from Monsanto at various levels of pulse excitation. This diode was then irradiated with 2 MeV electrons and the spectrum and the change in the laser threshold detected. Several examples are given to indicate the type of data that have been obtained. Figure 7.8 illustrates the effect of the level of excitation upon the spectrum of a diode irradiated with 5×10^{17} e/m^2. Similar spectra are observed for unirradiated diodes. Figure 7.9 presents spectra after irradiation with 5×10^{17}, 10^{18} and 2×10^{19} e/m^2. Although the excitation intensity is not the same in all cases, it appears that a genuine change in the spectra is resulting from the irradiation. This is further confirmed in Fig. 7.10. The spectral width of the spectrometer was 11 pm for these measurements. The laser action of the diode was quenched after irradiation with 10^{21} e/m^2. This diode was then annealed several times at temperatures between 375 and 433°C in an argon atmosphere. It was again found to lase with a threshold at about 37.5 A. The emission spectrum of the annealed diode is shown in Fig. 7.11. The

[†]The advice and assistance of Professor N. Holonyak, Department of Electrical Engineering, and of the graduate students in his group is gratefully acknowledged.

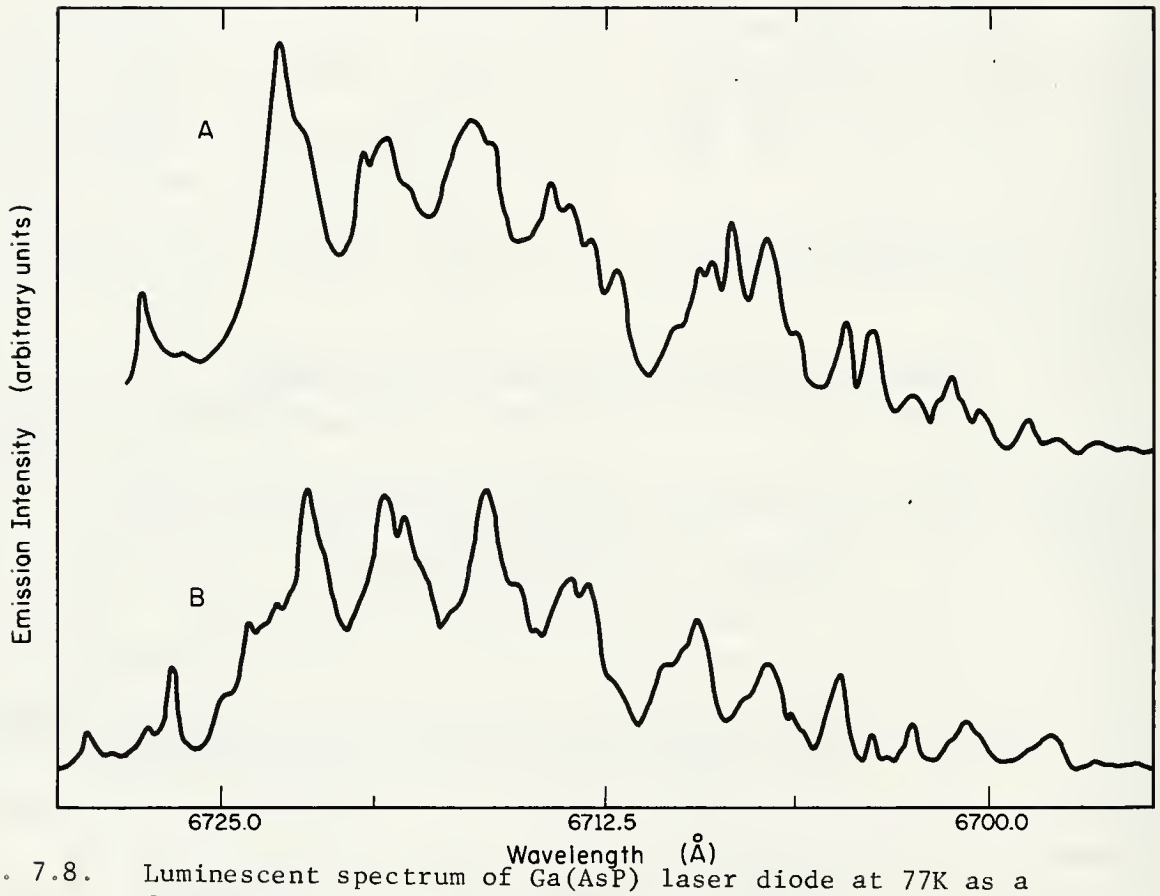


Fig. 7.8. Luminescent spectrum of Ga(AsP) laser diode at 77K as a function of excitation intensity: Curve A 44.5A Current Pulse, Curve B 38A Current Pulse. Scales for A and B have been shifted. Spectra taken after an irradiation of 5×10^{19} e/m^2.

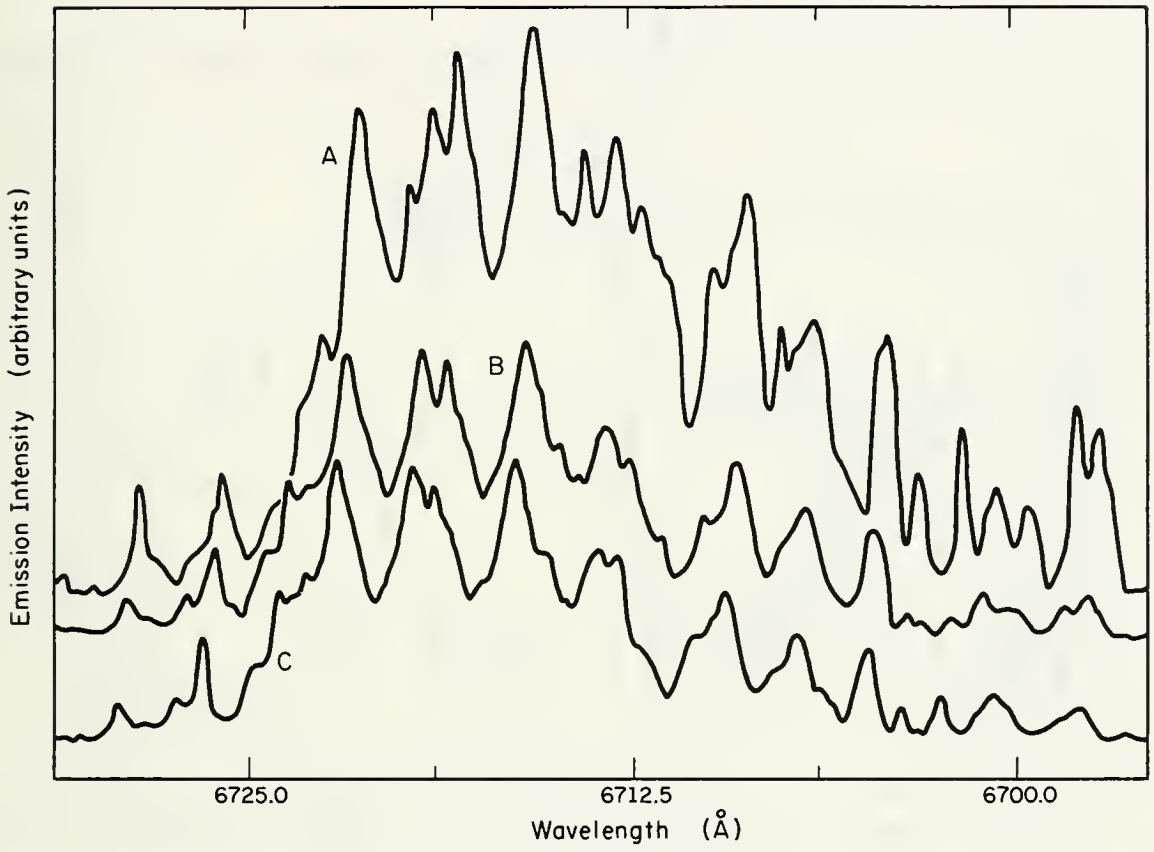


Fig. 7.9. Luminescent spectra of Ga(AsP) laser diode at 77K after irridiation with various fluxes of fast electrons. Curve A: 5×10^{19} e/m^2 38A Current excitation. Curve B: 10^{20} e/m^2 38A Current excitation. Curve C: 2×10^{21} e/m^2 33A Current excitation. The vertical scales for the three curves have been arbitrarily shifted.

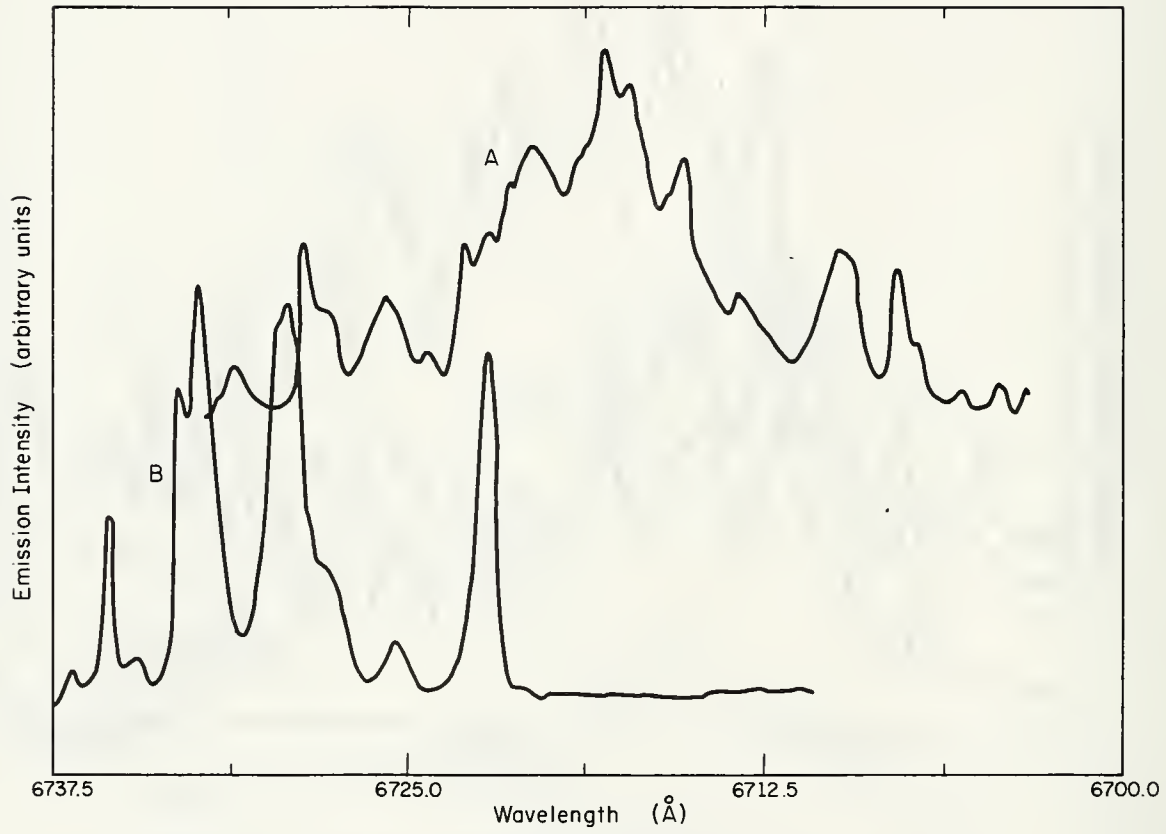


Fig. 7.10. Luminescent spectra of Ga(AsP) laser diode at 77 K after irradiation with different fluxes of fast electrons. Curve A: 1×10^{22} e/m^2 33 A Current excitation, Curve B: 5×10^{22} e/m^2 46 A Current excitation. The vertical scales for the two curves have been arbitrarily shifted.

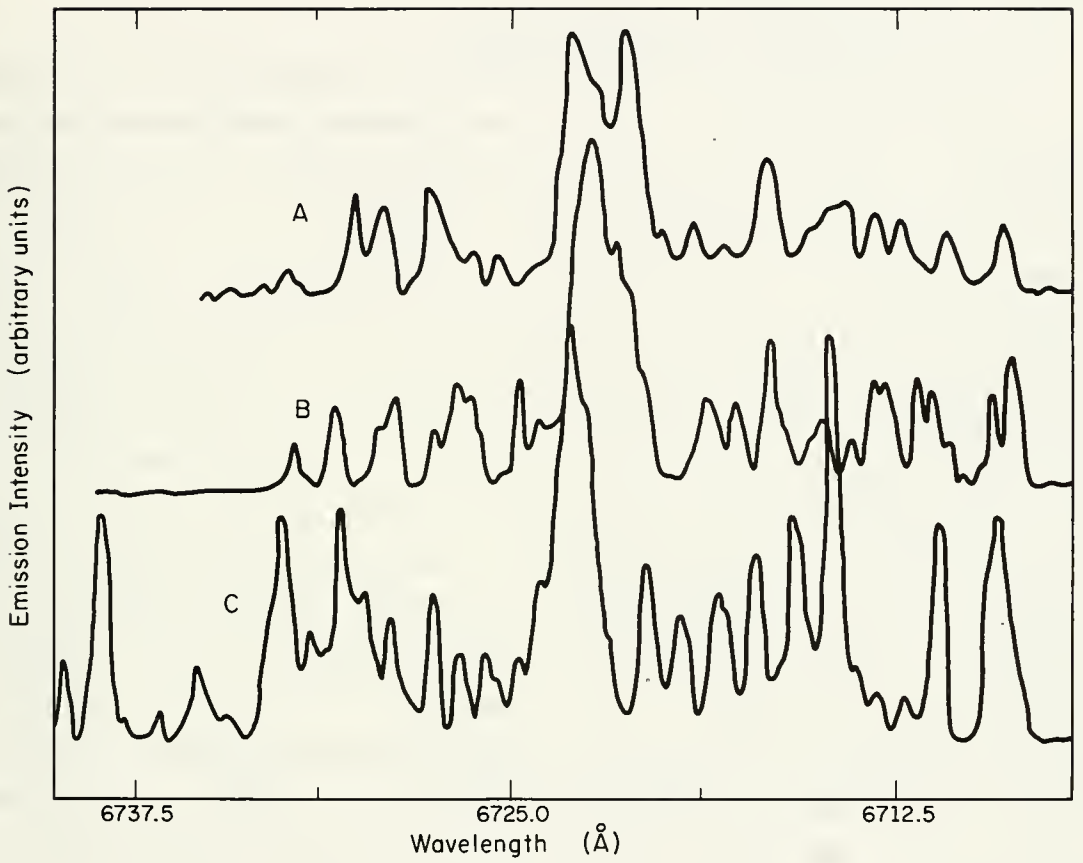


Fig. 7.11. Luminescent Spectra of Ga(AsP) laser diode at 77K after irradiation with 10^{23} e/m^2 and anneal in Argon atmosphere. Curve A: Annealed at 781 K. Curve B: Annealed at 748 K. Curve C: Annealed at 706 K. All measurements made at 49 A excitation. The vertical scales for the three curves have been arbitrarily shifted.

dependence of the threshold for laser action as a function of electron irradiation is shown in Fig. 7.12.

It is clear that the effects of the irradiation are quite complicated for a p-n diode. It is expected that the effects will be much less complicated for platelets and thus, the subsequent research effort has concentrated on CdSe platelets.

7.3.4. Laser Action in CdSe Platelets

Laser action can be readily induced in CdSe platelets by pumping with a Ga(AsP) laser [4,5]. It is evident, however, upon looking at the emission from a variety of platelets, that the position of the laser luminescence line is not always reproducible. A number of factors are apparently responsible for this, some of which are the geometry of the sample, the level of excitation, and perhaps even the purity of the material. Of particular interest, and the prime objective of this research, is the role of defects in the laser process. Before examining the influence of radiation-induced defects upon the laser action of these platelets, a detailed examination of the laser action prior to irradiation has been carried out on one platelet. The results of such a measurement on CdSe is shown at 77K in Fig. 7.13. Figure 7.13 gives the spectrum of the laser line at high resolution for four levels of excitation. The excitation levels were achieved by applying different voltage pulses to the Ga(AsP) laser diode driver, the voltages for each curve being given in the caption. Several observations are relevant. At lowest excitation, the regular mode pattern of the laser is evident. At highest excitation, the mode pattern becomes less distinct and long-wavelength peaks appear which are not clearly part of the mode pattern. Finally, the position of particular well-defined modes are seen to shift to shorter wavelength with the increase in the level of excitation.

A preliminary observation of the time dependence of the luminescence of the principal peaks of Curve A of Fig. 7.13 has been made. It is found that the short-wavelength peaks generally occur earliest with the long-wavelength peaks occuring latest in the pulse.

Although these data suggest that several factors can influence the location and the spectral dependence of the laser line, it is not at all clear how the microscopic features of the transition are affected by such parameters as excitation intensity, or whether localized transitions, arising perhaps

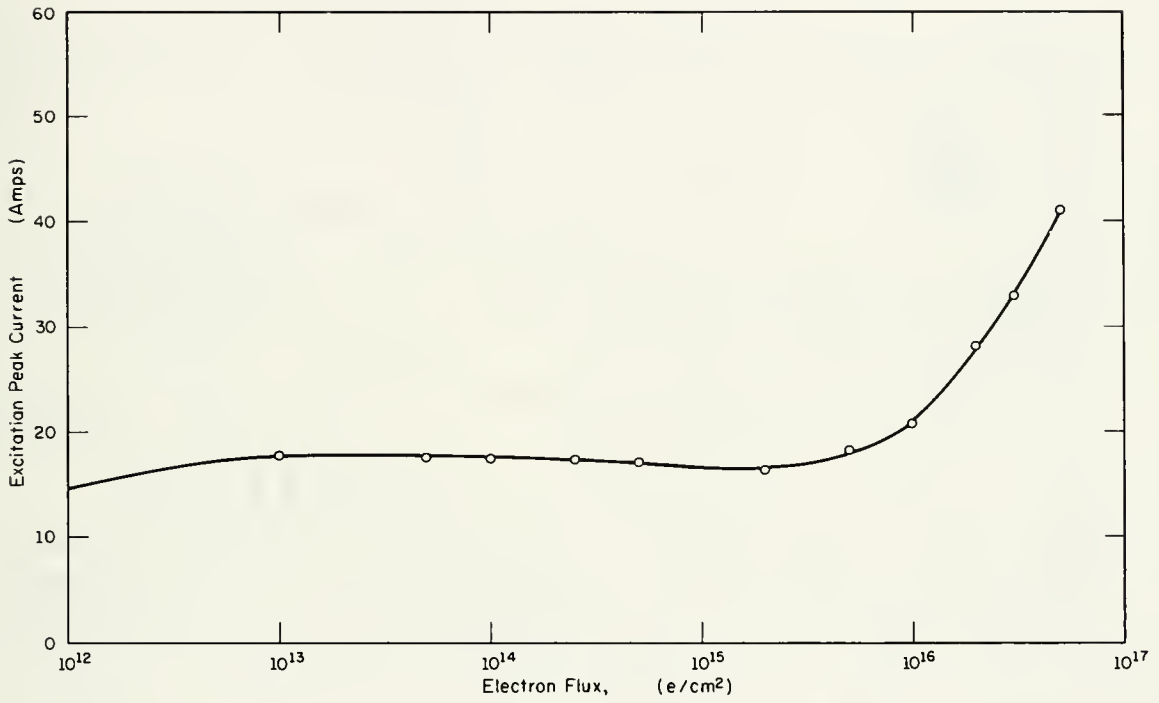


Fig. 7.12. Variation of threshold voltage for laser action of Ga(AsP) diode at 77 K as a function of 2 MeV electron irradiation.

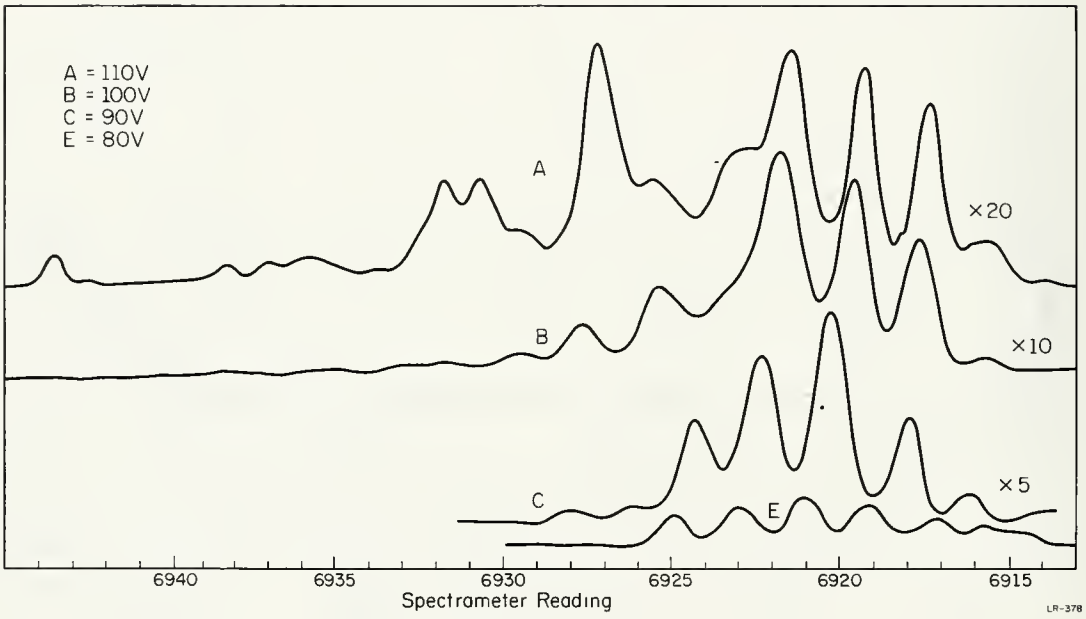


Fig. 7.13. Laser emission at 77 K from a CdSe platelet, optically pumped with Ga(AsP) laser diode. Curves A,B,C and E correspond to a driver voltage of 110, 100, 90 and 80 volts, respectively. Vertical scale factors given on each curve.

from defects, are involved in the laser transition. Irradiations are underway with these platelet and identical measurements to the preirradiation will be made.

7.4. Investigation of P-I-N Diodes as Infrared Detectors[†]

P-i-n devices fabricated from silicon containing dopants such as Co, Zn, and Au, exhibit current oscillations in the positive resistance region of their forward characteristics [6,7,8]. The frequency of the current oscillations and the photoconductive properties of these devices are strongly affected by radiation in the near infrared, giving rise to the possibility of using these devices as infrared detectors. It is believed that the current oscillations can occur whenever recombination centers exist whose cross section for electron and hold capture differ markedly, and when space-charge-limited current flow exists. It is likely, therefore, that oscillations can occur for many types of deep-level defects in many different materials [6,7,8]. This work is directed at achieving a better understanding of these phenomena, particularly of the trapping and recombination centers that give rise to the current oscillations.

A natural question arises as to whether defects other than impurities can operate to produce an oscillatory-current characteristic. Since irradiation with high-energy particles introduces a variety of structural defects having energy levels distributed throughout the forbidden gap, it seemed likely that irradiation could serve as a useful technique for introducing a variety of defects [9].

P^+-n-n^+ devices were fabricated from n-type, 140 Ωm (about $3\times10^{19}/m^3$) silicon. As fabricated, the devices exhibited typical I-V characteristics. These devices were then irradiated at 0^oC with approximately 10^{20} e/m^2 of 1 Mev [10]. After irradiation, current oscillations appeared in the positive-resistance region of the characteristics at temperatures below 250K.

The threshold oscillation frequency is a function of temperature and decreases as temperature decreases. Figure 7.14 shows frequency-vs-

---

[†]Drs. B. Streetman and N. Holonyak, Dept. of Electrical Engineering are investigating the mechanisms of current oscillations in p-i-n devices. The program reported here was carried out in conjunction with their program. Their advice and assistance are gratefully acknowledged.

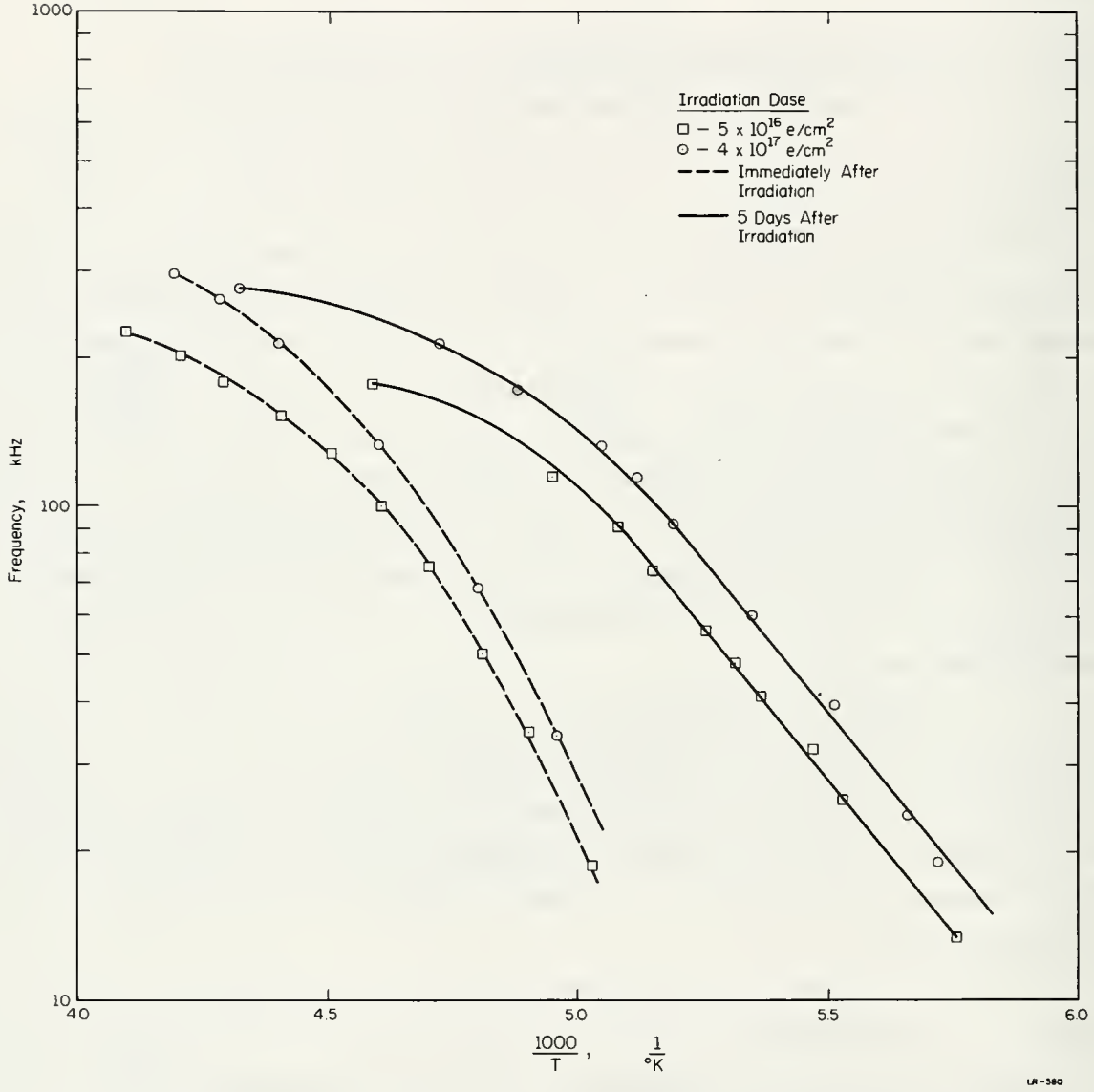


Fig. 7.14. Temperature dependence of threshold oscillation frequency and the effects of annealing for two devices with different irradiation doses.

reciprocal-temperature data on two devices that were subjected to different levels of irradiation. Curves were taken immediately after irradiation, and at five day intervals, while annealing at liquid-nitrogen temperatures. A large change in oscillation frequency was observed after the first five days of annealing. No significant change occurred for later measurements. The observed change in oscillation frequency suggests that such annealing studies may be used as a measure of the defect properties of p^+-n-n^+ devices. In particular, there is little information available on the motion of defects at 77 K. This study has also shown that the requirements for oscillation in p-i-n devices are met by structural defects created by irradiation with high-energy electrons as well as by impurity-induced traps.

Several devices were fabricated and irradiated to various flux levels between 10^{20} and 10^{21} e/m^2. The irradiation temperature was approximately 95 K. These devices were divided into two groups and allowed to anneal at temperatures of 200 and 273 K. Periodic frequency-*vs*-temperature measurements were made during the annealing period. Unfortunately, the annealing data on these devices did not show a definite pattern. The oscillation frequency ranged from 10 to 400 kHz for this group, and oscillations were observed to continue for approximately two weeks at the annealing temperatures.

Data have also been taken on threshold oscillation frequency *vs* irradiation level for devices fabricated from both pulled and float-zone silicon of different resistivity and different-area junctions. Although the plots of frequency *vs* irradiation flux appear to be quite complicated, the general trend is for an increase of frequency with flux, as one would expect for a continuing increase in the concentration of defects with irradiation flux. Figure 7.15 is presented as one example of data that are obtained for three devices fabricated from float-zone grown material.

The first evidence for oscillation at the low flux levels yielded a nonsinusoidal behavior, the frequency of which did not lie on the curves as plotted. All points as shown were taken from sinusoidal behavior. Subsequent annealing studied on these devices did not exhibit a definite pattern with time. It thus appears that parameters that are not being controlled are influencing the oscillatory behavior of these devices and that a better understanding of these factors is imperative. The irradiation has demonstrated the general validity of the concept that defects introduced in a variety of

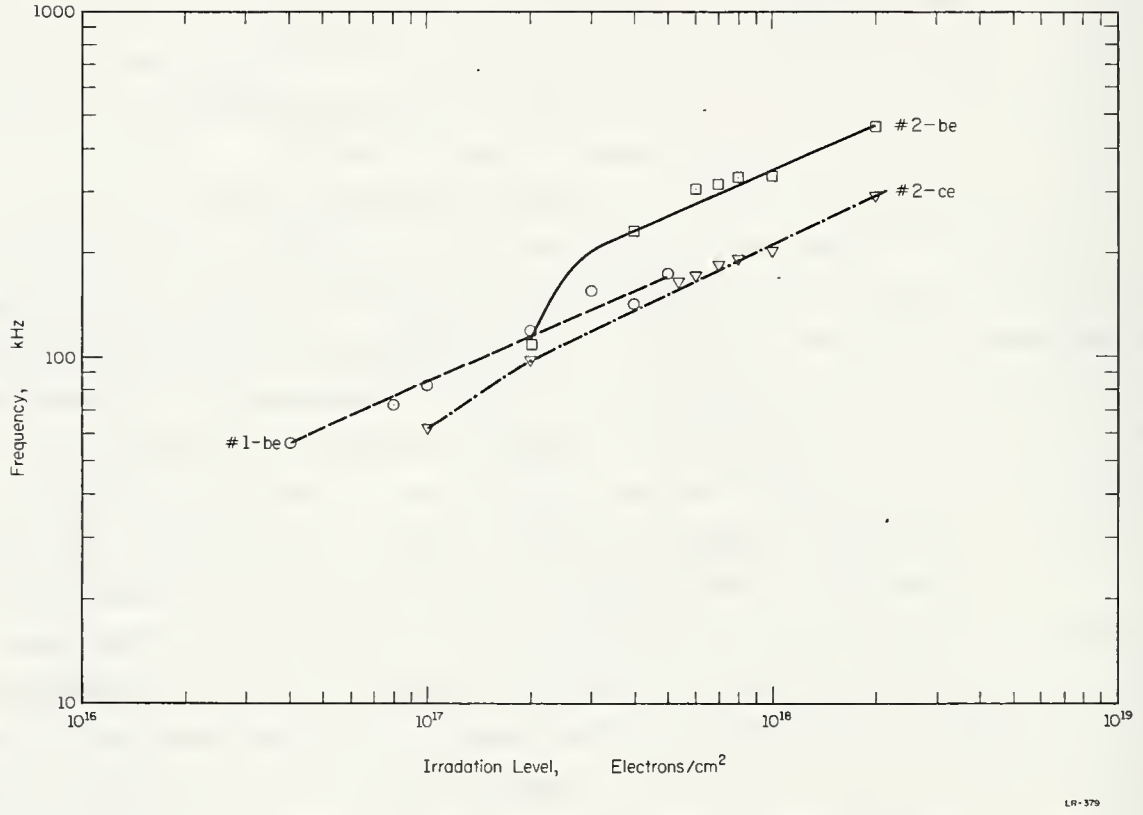


Fig. 7.15. Oscillation frequency vs. irradiation level for P-I-N diode fabricated from 1.0 Ωm n-type float zone silicon. Width of intrinsic region was 100μm.

ways can generate oscillations and that impurities are not essential for this phenomena.

7.5. Metal-Semiconductor Tunneling

7.5.1. Tunneling in P-Type Silicon

Tunneling measurements have been completed on MOS devices made from degenerate silicon. The data shown in Fig. 7.16 through 7.19 were obtained from samples cut from 4 boron doped crystals with impurity concentrations ranging from 1.8×10^{25} m^{-3} to 2×10^{26} m^{-3}. The interaction of tunneling electrons with optical and local-mode phonons is clearly evident in the data. Samples with the highest boron concentration, Fig. 7.16, exhibit phonon lineshapes that are qualitatively explained in terms of a many-body self-energy effect in the semiconductor electrode.[11]. Variations in the lineshape with boron concentration (Figs. 7.17, 7.18 and 7.19) lead us to the conclusion that both the electronic self-energy effect and inelastic phonon emission in the barrier must be taken into account in order to adequately describe the observed data.

Figure 7.16 shows di/dV and d^2i/dV^2 data from a sample with a boron concentration of 2×10^{26} m^{-3}. Positive bias corresponds to raising the Fermi level in the metal with respect to that in the semiconductor. The large positive d^2i/dV^2 peaks at positive and negative bias results from the interactions of tunneling electrons with silicon optical phonons of small wave vector. The positive bias peak occurs at 64.2 ± 0.4 mV which is 0.6 mV below the value obtained from Raman scattering for the optical phonon. This discrepancy is believed to be due to phonon dispersion effects. Since the peak gradually shifts with decreasing boron concentration to 64.5 ± 0.4 mV in the lowest-doped sample, this explanation would seem to be valid.

At positive bias the interaction with silicon optical phonons results in an increase in the conductance while at negative bias a conductance decrease is observed. The corresponding second derivative structure is approximately symmetrical about zero bias. Inelastic phonon emission in the barrier would have resulted in an increase in conductance in both bias directions and hence antisymmetric peaks in second derivative. This inelastic tunneling process can therefore be ruled out as the predominant mechanism in this case.

Duke and Davis [12] have calculated the electronic self energies in degenerate semiconductors due to electron interactions with optical phonons;

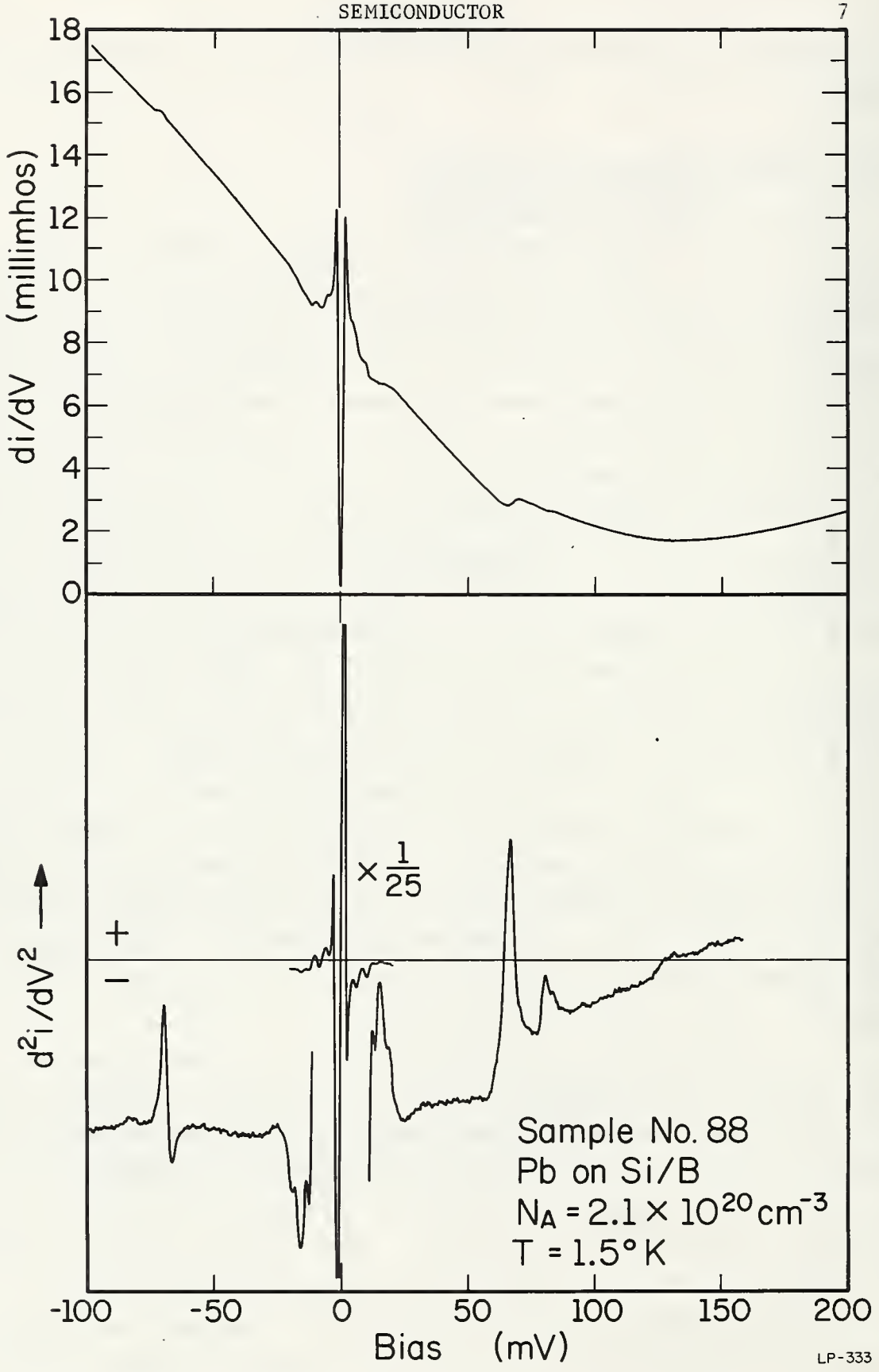


Fig. 7.16. di/dV and d^2i/dV^2 for indium-silicon junction with silicon doped to $N_A = 2.1 \times 10^{26}/m^3$.

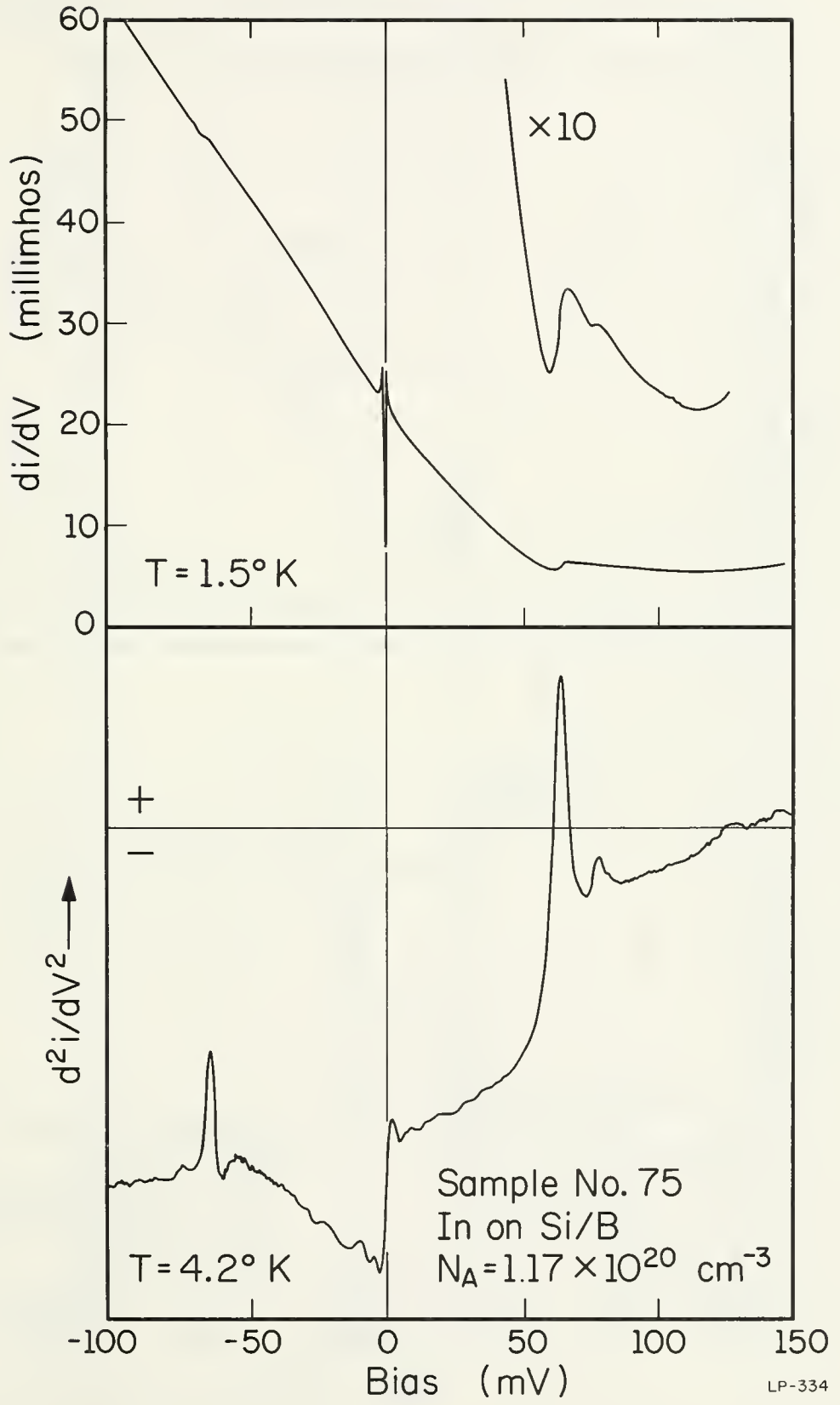


Fig. 7.17. di/dV and d^2i/dV^2 for indium-silicon junction with silicon doped to $N_A = 1.17 \times 10^{26}/m^3$.

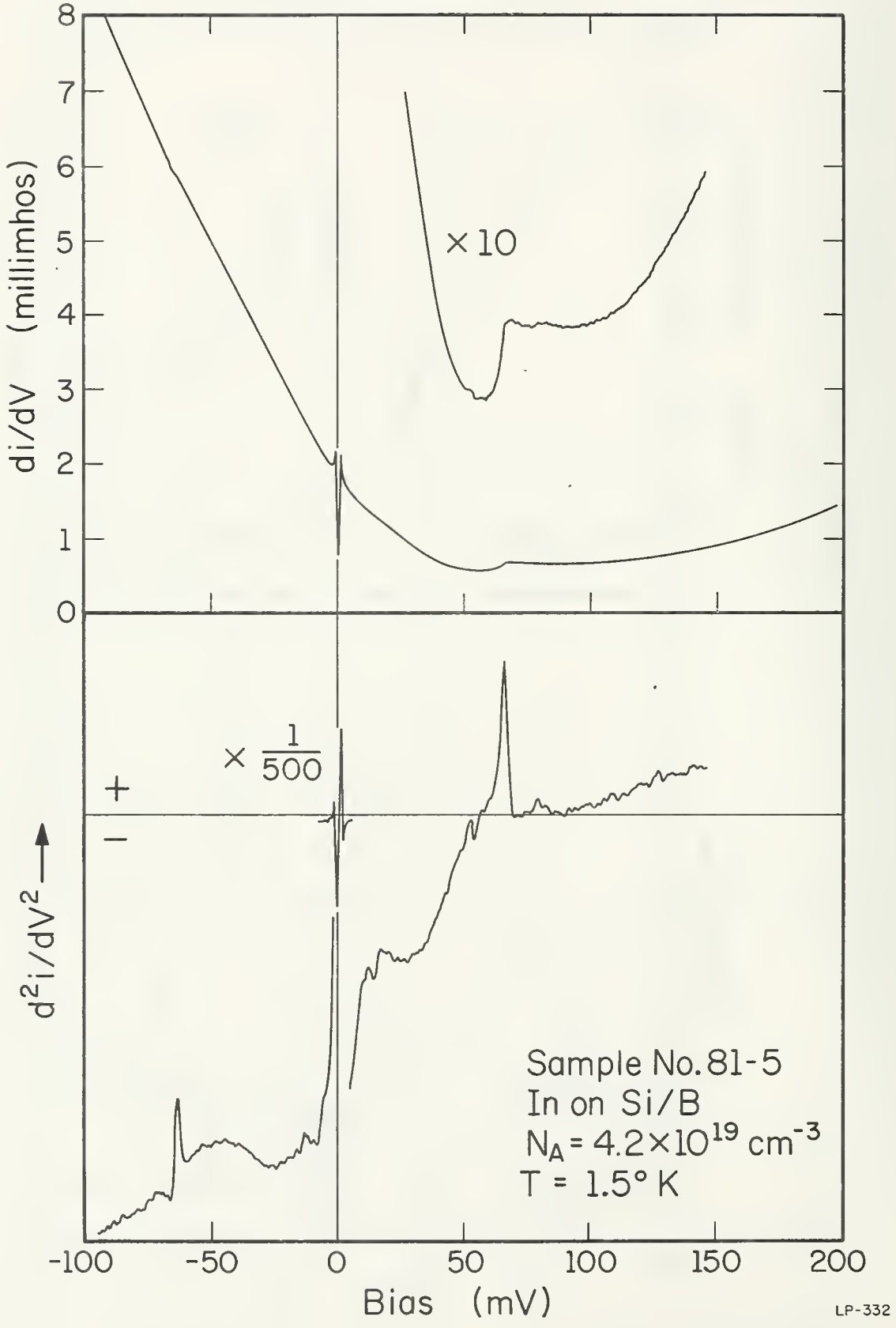


Fig. 7.18. di/dV and d^2i/dV^2 for indium-silicon junction with silicon doped to $N_A = 4.2 \times 10^{25}/m^3$.

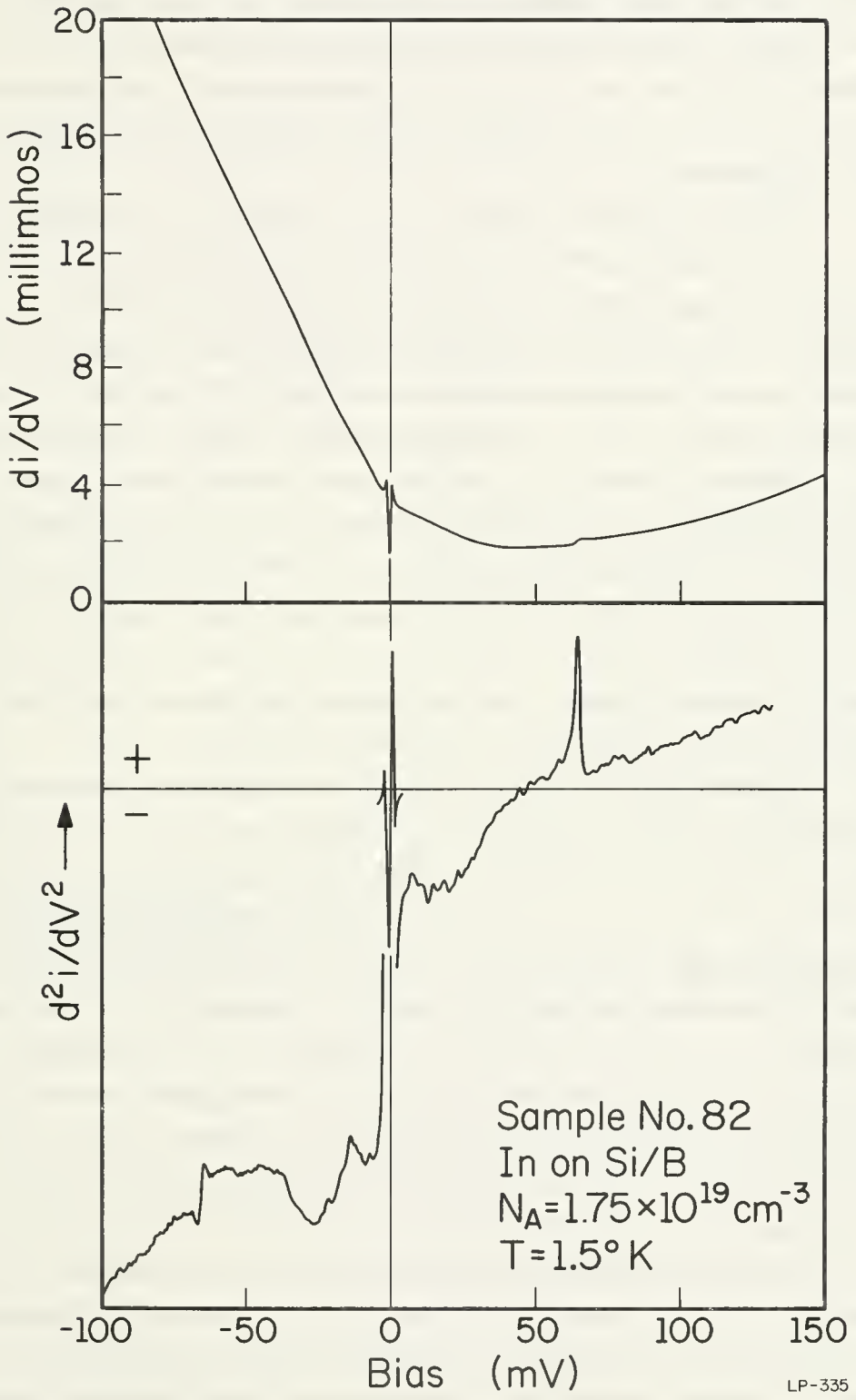


Fig. 7.19. di/dV and d^2i/dV^2 for indium-silicon junction with silicon doped to $N_A = 1.75 \times 10^{25}/m^3$.

and the effects of such many-body interactions on the tunneling conductance have been evaluated numerically and compared with these experimental data. The theoretical lineshapes are in qualitative agreement with experiment. The symmetrical d^2i/dV^2 structure is thus interpreted as being the result of an electronic self-energy effect due to optical phonon interactions in the semiconductor electrode.

Separated from the optical phonon peaks by about 15 mV is another pair of positive peaks. This structure has been attributed to interactions with boron local-mode vibrations. The origin of this structure is clearly established, not only by the energy of the peaks, but by their strong dependence upon boron impurity concentration and the details of the splitting of the structure at positive bias. Infrared-adsorption measurements on boron-doped silicon attribute peaks at 76.9 and 79.9 meV to local mode absorption due to isolated B^{11} and B^{10} isotopes [13]. The natural isotopic abundances are 80% B^{11} and 20% B^{10}. The relative heights of the two resolved peaks in forward bias, the separation of the two peaks, and the absolute energy of the peaks are all in agreement with the infrared-absorption data. The symmetric nature of this structure would again lead us to believe that we are observing the result of a many-body effect in the silicon electrode.

Figures 7.17, 7.18 and 7.19 show tunneling data from samples with progressively decreasing boron content. The boron local-mode structure grows weaker as the boron concentration is decreased and fades entirely at the lowest doping. While the positive bias optical phonon lineshape does not change significantly in these figures, the reverse bias d^2i/dV^2 structure undergoes a continuous change from a positive peak at the highest doping level to a step-like structure as the lowest doping level. In contrast to the first three samples, the Fermi degeneracy of the silicon for $N_A=1.8\times10^{25}$ m^{-3}, (approximately 45 meV), is less than the optical phonon energy; hence, at a positive bias equal to the phonon energy, electrons tunneling from states near the Fermi level of the metal electrode can do so only if they emit a phonon, since these electrons are tunneling into the forbidden gap of the semiconductor. We now have a situation in which inelastic phonon emission must be the dominant mechanism. That the many-body effect must still be considered, however, is implied by the reverse-bias lineshape. Similar lineshapes have been observed in heavily doped p-type Ge MOS junctions. Neither

mechanism alone predicts such a lineshape, and we feel that more precise calculations, including both the electronic self-energy effect of the optical phonon coupling and inelastic phonon emission in the barrier would be required to explain this latter lineshape and the variations in lineshape with boron concentration that we have observed in our data.

Preliminary measurements have been made on silicon samples of lower doping than presented in Figs. 7.16 through 7.19 with $N_A = 6.5 \times 10^{24}$ m^{-3}. On these units, the vacuum cleavage was made in a way that allowed some contamination of the cleavage plane; not enough for a insulating barrier to form, but enough to alter the surface-state density, producing a higher tunneling barrier and hence measurable resistances. While the superconducting structure of the metal electrode was not seen, a magnetic-field-dependent conductance minimum was observed at zero bias, as seen in Fig. 7.20. Also the zero-bias conductance minimum was found to be strongly temperature dependent, Fig. 7.21, and is thought to be the result of a two-step tunneling process. At positive bias, for instance, an electron tunnels into a localized state in the depletion region of the semiconductor and then tunnels into the valence band. Measurements of the temperature dependence over a wider temperature range together with a check of the variation with doping are planned. These experiments should make it possible to clearly establish the origin of the structure. If it is shown to result from localized states, this technique will be used to examine localized defects introduced by radiation.

7.5.2 Junction Fabrication

In the previous progress report [14], extensive results were reported on metal-semiconductor contacts on n-type Ge, p-and n-type GaAs, and p-and n-type Si. Extensions of this work will require measurements of junctions in magnetic fields and perhaps under high pressure. For either case, another method of making ohmic contact to the evaporated metal dots, the metal parts of the metal-semiconductor junction, is needed.

In the previous work, contact was made by lowering a spring-loaded, pointed indium wire onto the metal dot. This method obviously suffers from a lack of mechanical stability. A new method of making contact is now under development. Preliminary results indicate that the new contacts are much more stable and that the junctions do not deteriorate after a month of storage.

The procedure utilizes positive photoresist. After the metal dots

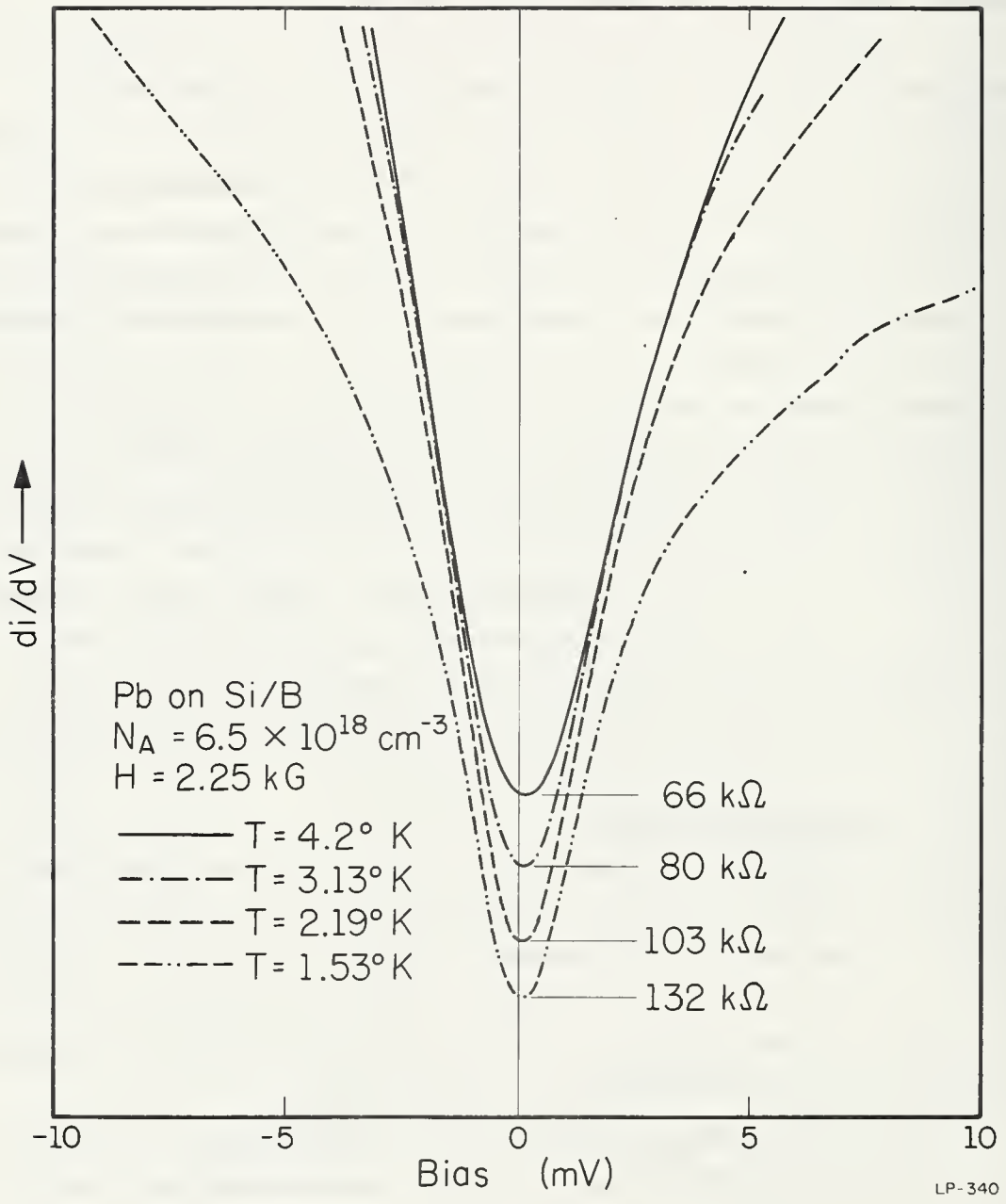


Fig. 7.20. Temperature dependence of the conductance vs. applied bias for a lower doped sample containing $N_A = 6.5 \times 10^{18}/m^3$.

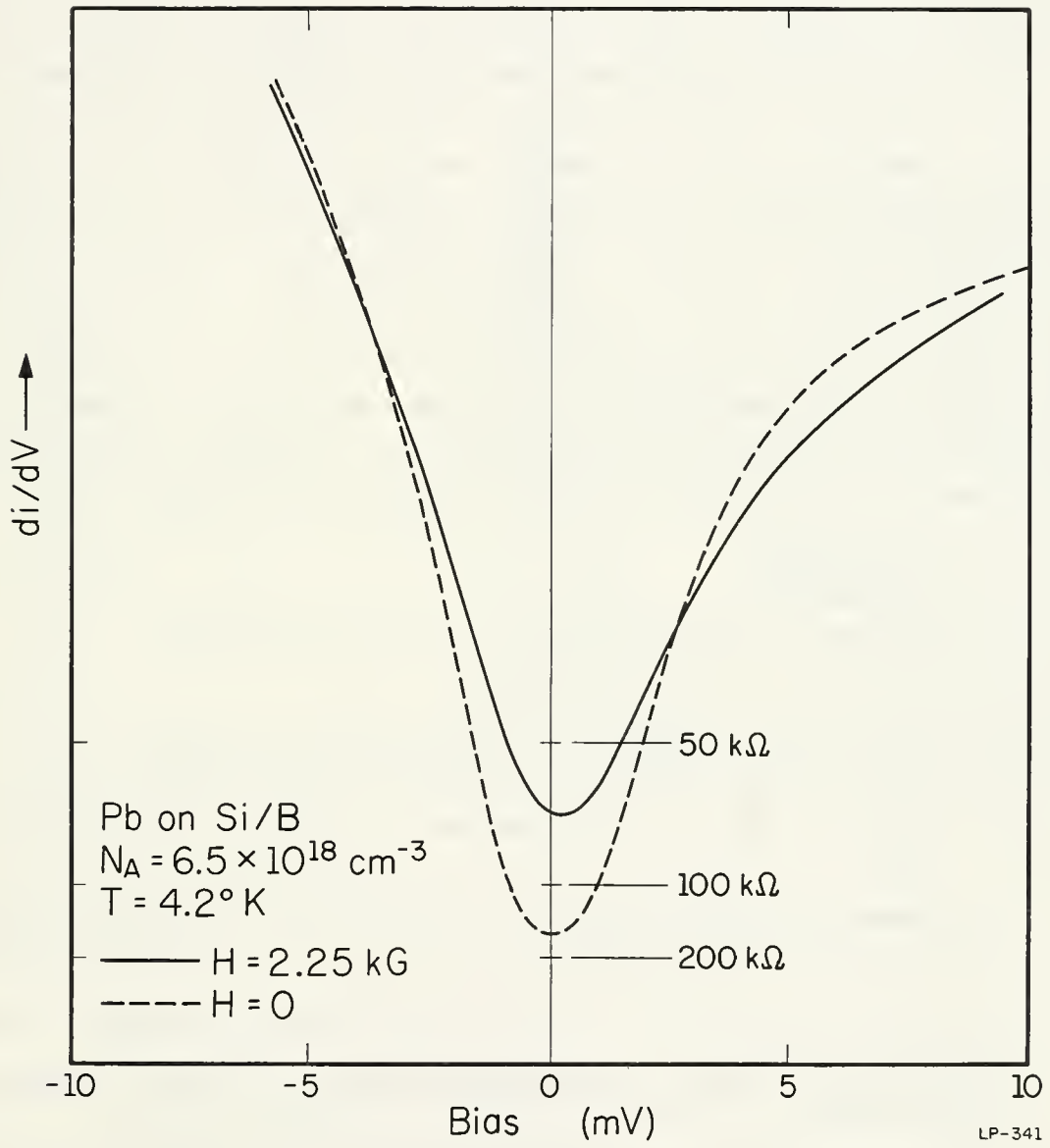


Fig. 7.21. Magnetic Field dependence of the conductance vs applied bias for sample shown in Fig. 7.13.

are evaporated onto the semiconductor, the semiconductor is covered with positive photoresist. This film is an insulator which, when exposed to light and developed in a special solution, will dissolve in the exposed area. The photoresist is exposed to light from a 100-W mercury lamp, focused with a metallurgical microscope directly onto a metal dot, and developed, exposing the metal dot underneath. Thick lead lines are then evaporated across the exposed dots. Contact to the lead lines can be achieved directly with solder.

Figure 7.22 shows d^2i/dV^2 of a metal-semiconductor junction of air cleaved p-type Si and a 0.2-mm diameter lead dot to which contact was made with the photoresist technique. The shape and position of the structure is identical to previous results using direct indium point contact. The position of the phonon peaks at 65.0 meV in forward bias and 65.7 meV in reverse bias indicates that any series resistence introduced by the photoresist is less than 0.1 Ω. The presence of the superconducting energy gap at zero bias indicates that tunneling is the current-carrying mechanism and furthermore, that the photoresist does not interfere with the superconductivity of the metal dot.

7.6. References

1. R.J. Spry and W.D. Compton, Phys. Rev. 175, 1010 (1968).

2. I. Balshev, Phys. Rev. 143,636 (1966).

3. W.A. Runciman, Proc. Phys. Soc. (London) 86, 629 (1965).

4. N. Holonyak, M. Sirkis, G. Stillman and M. Johnson, Proc. IEEE 54, 1068 (1966).

5. G. Stillman, M. Sirkis, J. Rossi, M. Johnson and N. Holonyak, Applied Phys. Letters 9,269 (1966).

6. N. Holonyak, Jr. and S.F. Bevacqua, Applied Phys. Letters 2, 71 (1963).

7. J.S. Moore, N. Holonyak, Jr. and M.D. Sirkis, Solid State Electron 10, 823 (1967).

8. B.G. Streetman, M.M. Blouke, and N. Holonyak, Jr. Applied Phys. Letters 11, 200 (1967).

9. R.A. Hewes, J. Applied Physics 39, 4106 (1968).

10. B.G. Streetman, N. Holonyak, Jr., H.V. Krone, and W.Dale Compton, Applied Phys. Letters 14, p. 63 (1969).

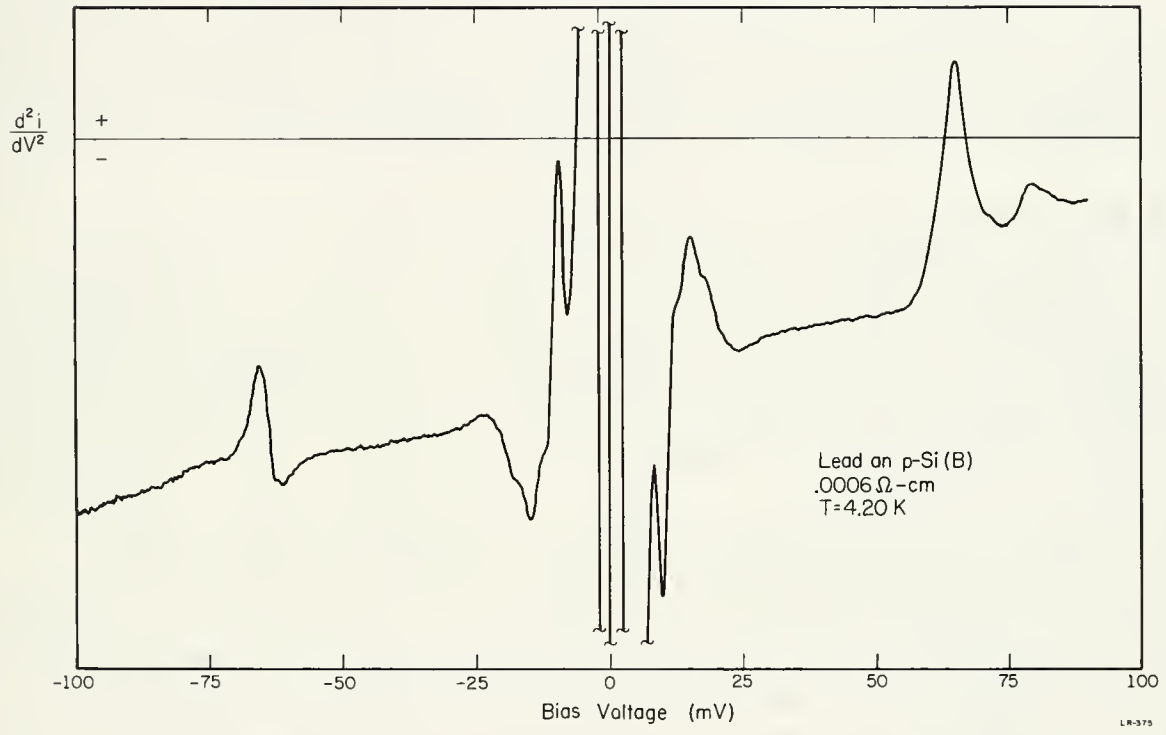


Fig. 7.22. d^2i/dV^2 for a lead-silicon junction with silicon doped to $N_A = 2.1 \times 10^{26}/m^3$. T = 4.2 K. Junction fabricated using positive photoresist technique.

11. E.L. Wolf, Phys. Rev. Letters 20, 204 (1968).

12. L.C. Davis and C.B. Duke. "Tunneling Measurement of Electronic Self Energies Due to Electron (Hole) Interactions with Optical Phonons in Semiconductors," to be published.

13. M. Balkanski and W. Nazarewicz, J. Phys. Chem. Solids 27, 671 (1965).

14. CSL Progress Report, March 1, 1968 through August 31, 1968.

H. G. Slottow
J. Bouknight
G. Crawford
L. Hedges
K. Kelley
J. Knoke
D. Lee
V. Metze
E. Neff
R. Resch
J. Stifle
R. Trogdon

8.1. Introduction

The Computer Group is responsible for the management, maintainence, and development of the CSL computer facility, and for the design of custom digital equipment and programs. The following sections discuss work in both hardware and software. The results on half-tone images (sections 8.6 and 8.7) have come from an unusual collaboration in CSL, and they warrant additional comment.

For about five years, an important part of the CSL computer program has been the development of techniques for displaying, graphically, the information generated by a digital computer. The current computer-controlled cathode-ray-tube system preserves the integrity of about 2000 lines in both x and y, and since it is driven by 12-bit digital-to-analogue converters, the quality of the images is high. Four programmable levels of intensity were provided in the original equipment.

This system was designed by engineers primarily for the display of engineering and scientific data. The quality of the system, however, and its potential for development attracted the interest of an architect (Professor R. Resch) who is interested in the study of new three-dimensional structures. The collaboration of Professor Resch, as a member of the group, with programmers and engineers has led to the development of hardware to provide 256 levels of intensity (section 8.7) and to the design of programs which exploit this capability and which also suppress hidden surfaces (section 8.6). It is believed that the quality of this system represents the state of the art in this area.

8.2 CDC 1604 Computer

Period:		August 30, 1968 to June 20, 1969
Total Running Time:		4161.8 hr
Average Per Day (7 day week)		14.25 hr
Operational Time:	97.42%	4054.3 hr

†This work supported by the Joint Services Electronics Program (U.S. Army, U. S. Navy, and U.S. Air Force) under Contract No. DAAB-07-67-C-0199.

Preventive Maintenance Time:	2.58%	107.5 hr
Emergency Maintenance Time:	0.00%	0.00 hr

L. Hedges
E. Neff

8.3. Systems Programming

A package of machine-language arithmetic subroutines was added to the ILLAR system master tape to facilitate the use of double precision and complex variables in FORTRAN programs. Since the compiler permits mixed-mode arithmetic, the package includes routines for each of the four operations--add, subtract, multiply, and divide--for any pair of operands, regardless of type. Also a complete repertoire of expontential routines of the type A**B, where operands A and B can be any mixture of types, is included in the package. Subroutines upon which these are dependent included double precision and complex exponentials and logarithms, as well as angle and magnitude of a complex number. Another package of routines for multiple-precision arithmetic was also provided, so that computations beyond double precision are now feasible.

The CSL-FORTRAN provides special codechecking features called for by the following control cards in a FORTRAN program:

1. CODECHECK
2. TRACE
3. BRANCH
4. LIMIT

The first of these activates the other three, the other three being individual codecheck feature controls. When the compiler reads one of these cards, the mode of compiling is extended to include calls to specially named routines to handle the condition. Although it was intended that these routines be user supplied, so that the user could do what he wanted with them, a set of routines to handle these features was written and put on the auxiliary master tape. These routines, which will be loaded when the user does not supply his own, have proven to be quite useful in the debugging of large FORTRAN programs. The names of the programs and their functions are as follows:

BRANCH--prints statement numbers of the program as they are executed.

TRACE--prints the name of the program as it is entered and as it is exited.

LIMIT--if a storage violation occurs, prints the value of the index, the size of the dimension, and the location of the violation.

In addition, several features auxiliary to these codecheck features have been added to the compiler, including the ability to specify printouts that are done only if the codechecking features are operative and the ability to turn the features "on" and "off" under program control.

The new RECURSIVE feature which was added to the FORTRAN compiler is to our knowledge unique among FORTRAN compilers. This feature allows FORTRAN subprograms to call themselves recursively, a capability that previously existed only in such languages as ALGOL, LISP, and the like. The feature was written to facilitate the writing of the hidden-surfaces algorithm. A package of routines CSLPSHPP was written to handle all saving and restoring of indexes and argument locations and to take care of the updating of pointers to data.

The FORTRAN compiler has been extended to include a "TITLE" feature which enables the programmer to specify, just once, a set of cards that are to appear in each of several programs on the same tape. Use of this feature is indicated when several programs must each have identical DIMENSION and COMMON declarations.

Among the other improvements to the FORTRAN compiler in this reporting period are:

use of statement numbers as arguments to subroutines,
multiprogramming-type FORTRAN programs,
DATA statement in FORTRAN,
COMMENT statement for printer comments at execution,
$$ for listing comments in FORTRAN programs,
INDEX for designation of index registers by name,
updated driver routine for the compiler and assembler.

A new system routine has been written for the cathode-ray-tube display unit which employs light-pen control via a tracking cross and up to 16 function "light buttons." The tracking cross may be made to follow the light pen or act as an increment generator (joystick).

Other system work, included routine work on the assembler, loader, and the various system service routines. The rate of discovering bugs and making appropriate changes in the system has dwindled to about one a month.

K. Kelley

8.4. Teletype

We have installed two units of the Teletype Corporation Model 33 TBE Standard Duty Send-Receive Page Printer. These units do not have the phone-line hookup but do have the tape perforator and reader. The intended use is for typing and producing paper-tape input to the CDC-1604 computer. System routines have been revised to enable input and output via teletype code. The teletype as used with the ILLIAR system has been assigned an equipment designator and an appropriate pseudoselection code. The EDIT routine determines for itself whether it is looking at teletype tape or flexowriter tape.

K. Kelley

8.5. Manuals

The manual for CSL-FORTRAN is complete and in the process of being printed. A catalog of subroutines along with descriptions of their use in the system is being written. This manual will be issued in sections by order of potential usefulness tc system users. The table of contents and the first chapter are in the hands of the typist. Other sections will follow shortly.

An operations manual also has been written and is in the final stages of preparation for printing. The manual for the CSLx system is also being revised to a permanent version, and the ILLIAR manual is being revised. The completion of these manuals brings the development of the ILLIAR system tc completion.

K. Kelley

8.6. Computer-Produced Half-Tone Images

The effort to produce half-tone images on the display console has led to improvements in both hardware and software techniques for computer displays. Improvements to the hardware are discussed in section 8.7.

8.6.1. Original Algorithm

The original algorithm which was used to produce half-tone images grew out of the work of John Warnock at the University of Utah[1]. This algorithm was revised and wedded into a three-step procedure:

(1) define a structure in FORTRAN language and produce a "wire-frame"

picture of it

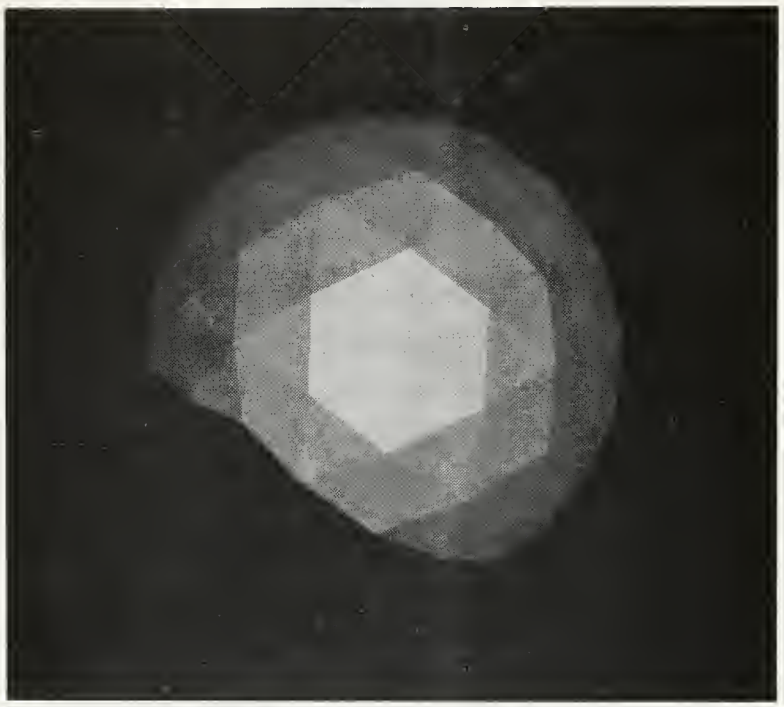
(2) produce a set of "key squares" from the "wire-frame" data using the revised Warnock algorithm

(3) produce a half-tone photograph from the "key square" data

A structure is defined by a FORTRAN program which calls upon a package of routines controlled by a driver and interpreter routine called ING. This package, developed by Professor R. Resch, is currently used in teaching applications of computer graphics to architectural and industrial-design students. The package provides such operations as defining points, lines, and figures in space, rotation and translation of the figures in space, and movements of the observer. The package has been extended to include the definition of planar polygons with specified coefficients of reflectivity for the surface. In the original application, the output from these programs was in the form of perspective photographs on CalComp plots of wire-frame structures. In this application the data for a set of planar polygons is written on magnetic tape.

In the second step of the procedure, this planar polygon data is read from magnetic tape and the algorithm is used to determine which faces are visible. The method is essentially a subdividing of the total view plane into successively smaller subsquares which are considered to be subpictures. The basic question asked at each step of the subdivision process is whether the resultant subpicture is simple enough to compute. The subdivision process terminates when a subpicture of unit (raster) size is reached. These will be unit squares along the visible edges in the picture. When one of these "key squares" is located its position is saved on magnetic tape, along with the number of the polygon which is visible at the upper left corner of the square. The total set of such key squares data defines a half-tone picture.

The third step of the procedure is to produce a half-tone picture from the key squares, which simply locate the positions on the raster where the beam passes from one polygon to another and thus necessitates changes in the intensity of the beam. For the purposes of these pictures, the scope hardware was modified so that 256 levels of intensity are available. The PIXSCANR routine generates the commands to the scope to control the beam between key squares. In this way a complex picture can be produced with a minimum number of data points. Two examples are shown in Fig. 8.1.



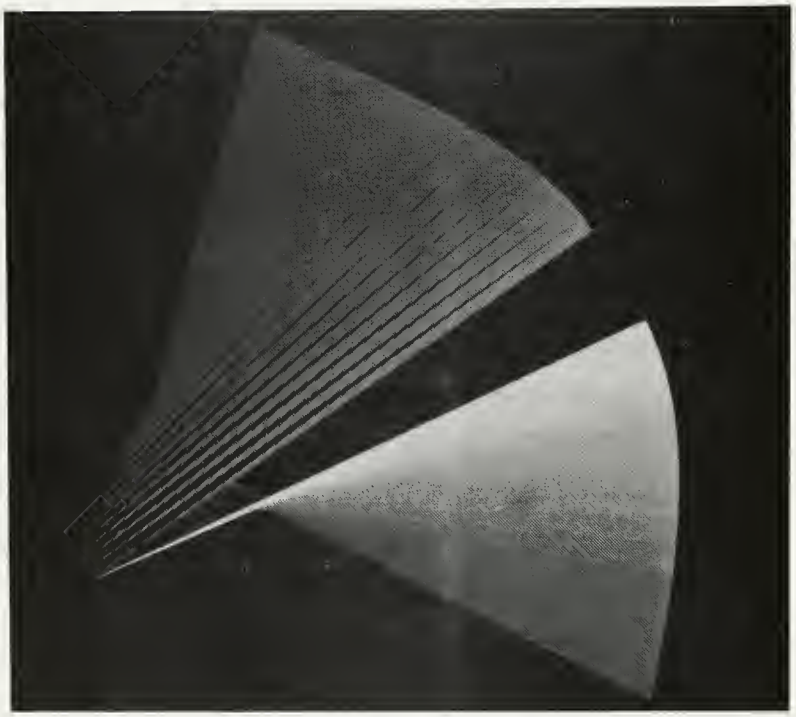


Fig. 8.1. Two examples of computer-generated half-tone pictures. The original quality (1000x1000) is degraded by the publishing process.

Most of the codechecking of the routines was done using, as test figures, an A-frame structure consisting of 17 polygons and an approximation to a torus consisting of 225 polygons. As we progressed through three revisions of the code and the algorithm, the times required for the computation reduced from the order of 4 and 17 minutes, respectively, for the second revision. We found the time required to be about porportional to the total of the lengths of the visible lines, for a given resolution, more so than to the complexity of the figure. Based on our experience with this algorithm, a new mode was added to the scope and a completely new algorithm produced which has reduced computation times even more.

K. Kelley

8.6.2. Improved Algorithm

A new algorithm for computer generation of the half-tone representation of a three-dimensional universe represented with planar polygons was developed as a result of experience gained in the implementation of the Warnock algorithm. The new algorithm is a raster-scan development of the final picture based on the requirement that as one scans a picture from left to right, one needs to detect when the scan point only enters or leaves the various polygons of the figure to be able to reproduce a half-tone representation, including hidden surfaces and changes in reflectivity in each surface.

The new algorithm obtains speed increases over the CSL-implemented Warnock algorithm of 6-12 times. Projections show that a CDC 6600 computer probably could generate a picture for a 300-polygon, 1200-edge universe in 3-5 seconds, as opposed to 2 to 2-½ minutes on the CDC 1604.

During the next reporting period, a full report detailing the algorithm with examples will be issued. Further work in this area will be directed to extending the algorithm to the computation of shadows produced by light sources positioned away from the viewer.

J. Bouknight

8.7. Display

A variable-intensity digital-scan mode has been installed in the display. In this mode, data are plotted as a raster of points, each of which has an intensity specified by the program. To save memory space, the mode hardware has been designed such that the program need specify an intensity only at those addresses in the raster where the intensity changes.

The intensity data consist of 8-bit words which are delivered to an 8-bit DA converter. The output of the converter is then distorted to match the combined characteristics of the CRT and the flim being used. The distortion function is essentially a log function. The output of the distortion circuit is used to control the amplitude of the unblank pulse applied to the CRT.

J. Stifle

8.8. Development of the CSLx System

The final programming and usage manual for the CSLx system was begun during the reporting period. It will be completed by fall of 1969.

With the issuance of the programming manual, development of the CSLx system is complete.

J. Bouknight

8.9. New Movie-Production System

During the reporting period, the Mitchell 16-mm movie camera installation was completed and checked out. The basic operating package of subroutines for program generation of movie film was completed and checkout begun.

The new movie system retains the off-line mode of film generation used previously. Movie-film data are written on a magnetic tape by the user program. A system program then converts the user's movie-data tape into a final movie film.

More flexibility has been provided to the user due to the new camera configuration. He can use the following operations:

(1) movie film forward,

(2) movie film backward,

(3) hold film so several frames are overlapped,

(4) skip film.

Movies may also be made in color and up to four synchronous films may be exposed.

Further development of the movie system includes test shooting in color for establishment of system parameters. An on-line editing program for combining movie-data tapes into single long films by program is under study.

J. Bouknight

8.10. Variable-Format Data System

Development has begun on a system subroutine package which provides the facility for handling data on magnetic tape in a freely specified format.

The user will simply specify alphanumeric names for his data, and the system will handle all bookkeeping operations to insure proper handling.

Conservation of space on magnetic tape will be achieved by packing user-size data records into the maximum-size record allowed on magnetic tape.

Service routines such as sorts, merges, collators, and printout routines will also be developed.

J. Bouknight

J. Hansen

8.11. References

1. Warnock, J., "A Hidden Line Algorithm For Half-Tone Picture Representation," Tech Report 4-5, University of Utah, May 1968.

9. PLASMA DISPLAY[†]

D. L. Bitzer
H. G. Slottow
W. Coates
R. Johnson
W. Petty
J. Knoke
J. Schermerhorn
R. Trogdon

9.1. Introduction

The invention of the Plasma Display Panel at the Coordinated Science Laboratory was a response to the anticipated display needs of the PLATO computer-based education system. Responding directly to digital information, it stores and displays this information on a rectangular array of bistable gas-discharge cells. The essential structure is simple, and with development, it should be inexpensive to produce.

The present stage of development is indicated in Figs. 9.1 and 9.2 which show the appearance of both computer-generated information and optically projected images on a Plasma Display Panel. Figure 9.1 illustrates a future PLATO student terminal while Fig. 9.2, taken from a chemistry course written for PLATO, shows a nuclear magnetic-resonance spectrum, generated by the computer, against the background of a page in the PLATO lesson.

The Plasma Display Panel used for these figures is an example of the Owens-Illinois Digivue Panel developed and fabricated under license from the U. of I. Foundation. It is driven by a CDC-1604 digital computer through circuits developed at CSL, and the information is displayed on a 128X128 array.

The principles of the Plasma Display Panel have been described in several papers and in preceeding CSL progress reports. In this report, we describe studies in several areas of plasma display research. These include two "on" state operation, applications to memory, circuits, field calculations, stability theory, and color.

9.2. Operation in the two-on-state modes

In earlier CSL progress reports and CSL R-Reports, considerable discussion has been presented describing the theory and experiments which indicate that a Plasma Display Panel can be operated in a mode in which each of two

[†]This work supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) and by the Advanced Research Projects Agency of the Department of Defense and was monitored by the U.S. Army Electronics Command under Contract No. DAAB-07-67-C-0199, and in part by the Naval Air Development Center under Contract No. N2269-69-C-0013 and Owens-Illinois, Incorporated.

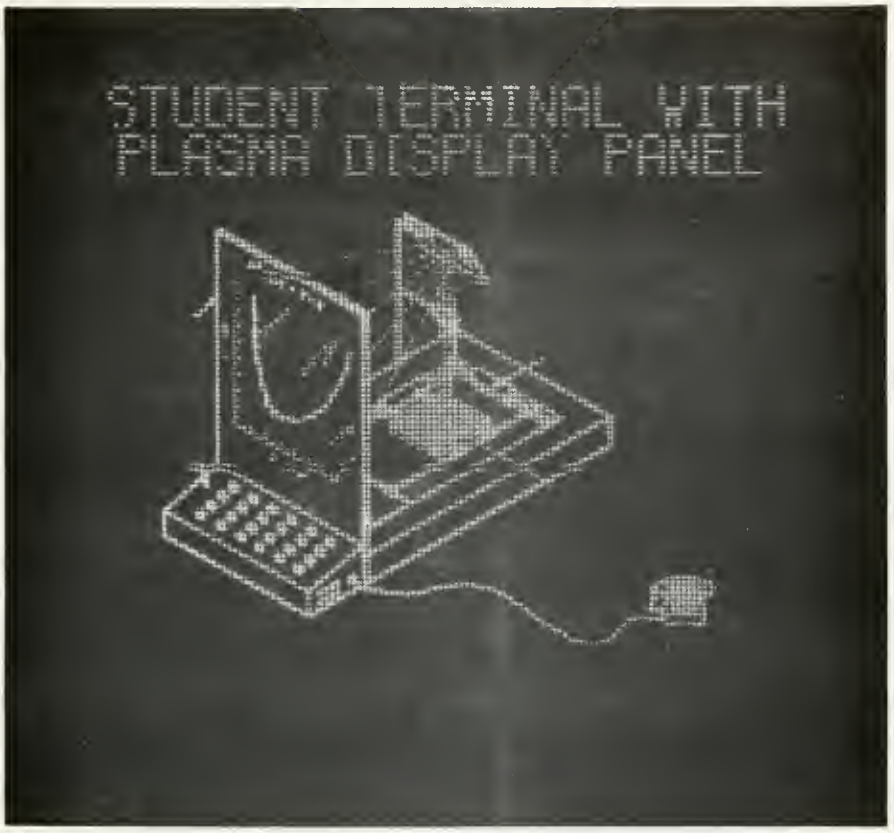


Fig. 9.1. Computer generated caption and projected image of terminal.

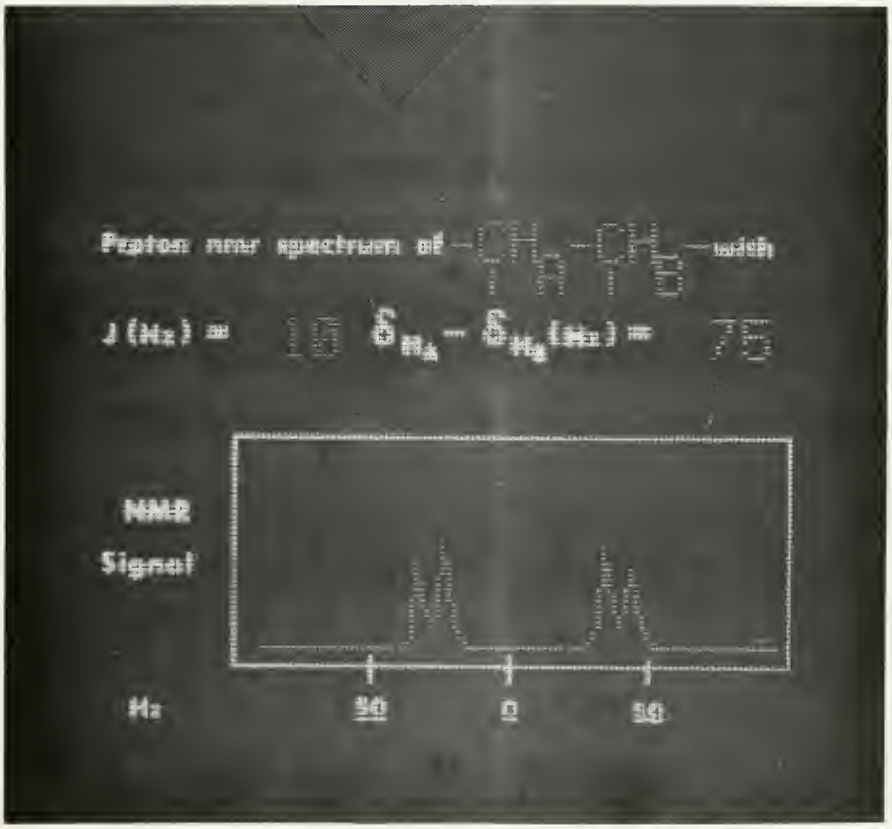


Fig. 9.2. Computer generated chemical formula and NMP spectrum and projection of PLATO lesson. These photographs show the superposition on a Plasma Display Panel, of computer-generated information and images projected from photographic slides. The quality of the projected images is degraded by the publishing process.

stable states assumed by a cell is characterized by a sequence of discharges, i.e., the two-on-state mode. The two stable states are electronically distinguishable from each other by the difference in the time of discharge occurrance, i.e., as phase information. A detailed discussion of the two-on-state mode of operation can be found in the CSL Report R-377 and in the CSL Progress Report of Sept. 1967 - Feb. 1968. Until recently, however, this effect was demonstrated only for single cells. During this reporting period, a sustain-voltage generator and address circuitry necessary for operating an entire panel in this mode were designed and tested on a 16X16-cell array in which each of the 256 cells was set in one or the other of the two stable on states by a coincident address-operation voltage signal. In the spot configuration of Fig. 9.3 the cells in the high intensity state (state A) are being sustained at six times the number of discharges per unit of time as those in the low intensity state (state B). Experiments performed on prototype 128X128 arrays show that discharge-rate ratios as high as 50:1 can be achieved. Plasma Display Panels operating in this mode can be used satisfactorily for visual data display purposes, since discharge-rate ratios as large as 50:1 provide optical contrast ratios between cells in opposite states which approach the optical contrast ratios of the standard on-off mode of operation. The bistability of each cell, however, remains intact.

Figure 9.4 illustrates a typical voltage waveform used to sustain a Plasma Display Panel which is operating in the two-on-state mode. Figure 9.5 shows the wall-voltage configuration and the light output of cells which are being sustained in each of the two states. The discharge-rate ratio between cells in each of the two states is determined by the ratio of the voltage waveform segments, labeled A and B, which are present in the sustain-voltage waveform. As indicated previously, the optical contrast ratio is directly related to the discharge-rate ratio. Figure 9.6 provides a detailed description of the sustain voltage waveform used on the 16X16-cell array shown in Fig. 9.3. (Note the optical contrast available with a discharge-rate ratio as small as 6:1). In the two-on-state mode of operation, as in the standard on-off mode, the transfer of a cell from one state to the other is accomplished with a single discharge of appropriate magnitude. The intensity of the discharge (i.e., the amount of charge produced) is accurately controlled so that the wall voltage of the selected cell is shifted to the wall voltage level of cells in the opposite

Fig. 9.3. 16X16-cell experimental plasma display panel operating in the two-on-state mode.

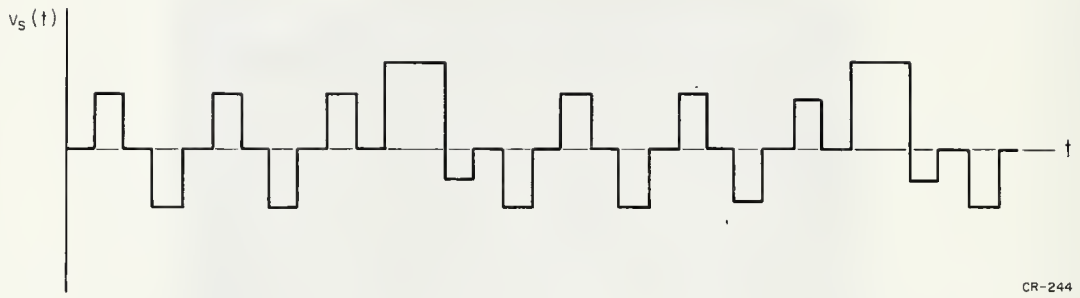


Fig. 9.4. A voltage waveform used to sustain a plasma display panel in the two-on-state mode.

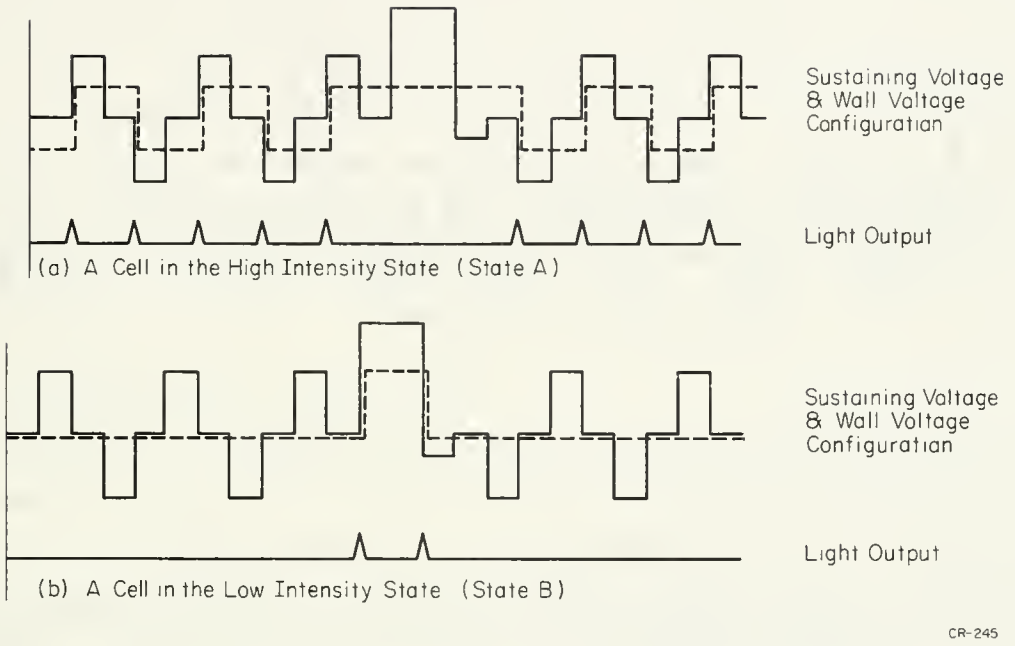


Fig. 9.5. Wall voltage, configuration, and light output for plasma display panel cells in each of the two stable "on" states.

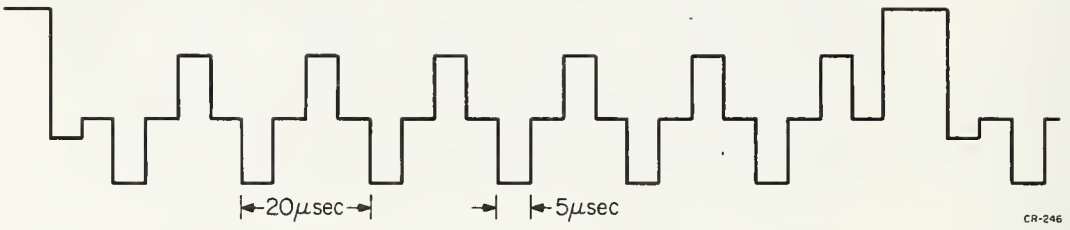


Fig. 9.6. Sustain voltage waveform applied to the experimental panel shown in Fig. 9.3.

stable state. This transfer is always accomplished without perturbing the wall voltages of any other cells in the array.

When a plasma panel is used as a digital display device, operation of the panel in the two-on-state mode can provide significant improvement in performance of the device. The improvements are related to the increased uniformity of the chemically and thermally dependent properties of the discharge mechanisms. Since there is now a guaranteed minimum, stable, discharge rate in all of the cells, the difficulties of starting a discharge sequence in a cell which has been off for some extended period of time (e.g., t greater than 1 millisecond) are alleviated. The minimum discharge frequency at which cells can be sustained is a property of the particular panel under consideration and is determined experimentally. That part of the sustain signal that shifts the cell to the dim state is sometimes called a conditioning signal.

9.3. Applications to memory design

Many of the circuit designs which were used to indicate the feasibility of the two-on-state mode of operation are now being applied to the continuing development of prototype random-access and associative-memory systems which will employ the Plasma Display Device as the bistable element array. The design and development of optical readout systems for these memories is the second major part of this effort. The high-density feasibility studies and the investigation of multiple-electrode techniques are two companion efforts which are related to the development of such memory systems as well as to the development of certain aspects of the visual display systems. A Plasma Display Panel consisting of 1600 cells at a linear density of 3.9/mm was constructed at the laboratory. This panel was operated with several gas mixtures; there were several qualitative indications which suggest that discharge sequence stability at these densities may be possible. However, no addressing of individual cells at this density has been accomplished to date. Several experimental panels have been constructed which involve multiple-electrode structures; these structures may lead to device-saving address schemes and may also be helpful in providing the inhibit function necessary in Plasma Display Panel memories. Efforts in this area are to be continued.

9.4. Circuits

During the previous reporting period, major emphasis has been placed upon the development of economical sustain-voltage generators and address systems for the digital-display applications. As a result of our increased understanding of the physical and electrical operating characteristics of the device, several design considerations have become more obvious. The Plasma Display Panel is a device which presents a very unusual electrical load to a sustain-voltage generator. When there are no cells in the panel that are undergoing pulse discharge sequences, the panel appears to be a pure capacitive load, e.g., a prototype 128X128 array represents approximately 1000 pF. However, when cells are undergoing discharge sequences, the panel appears to be a capacitance which has pulse-conduction current capability. This pulse-conduction current corresponds to the pulse discharges and is directly proportional to the number of cells which are supporting discharge sequences. A simple model of this type of electrical load is shown in Fig. 9.7. The design of a sustain voltage generator, therefore, requires that (1) the generator charge the capacity of the panel in an appropriate manner in order to provide the desired voltage waveform, and (2) the generator must be able to absorb the large pulses of conduction current without significant distortion of the voltage waveform. These requirements, in general, dictate that a very low-output-impedance voltage generator be used. As an example, Fig. 9.8a shows the voltage and current waveforms which are to be provided by a generator to a 128X128-cell array (i.e., 1000 pF) when no cells are undergoing discharge sequences. Figure 9.8b shows the voltage and current waveforms which are to be provided by the same generator when all of the cells are in the discharge mode (worst-case peak pulse current of about 1.5 A). An idealized model of a sustain-voltage generator which can provide the voltage waveforms described above is shown in Fig. 9.9. Because of the bridge configuration, the maximum voltage requirements for the switching components need only be one half the required peak-to-peak voltage of the sustain-voltage waveform. Sustain-voltage generators are now being developed at the laboratory which will satisfactorily approximate the characteristics of this ideal voltage generator. A sustain-voltage generator has been constructed which will operate a 128X128-cell array with less than 5% voltage-waveform distortion under peak load conditions, i.e., all cells in the on state. A similar low-impedance voltage generator has been constructed which will provide

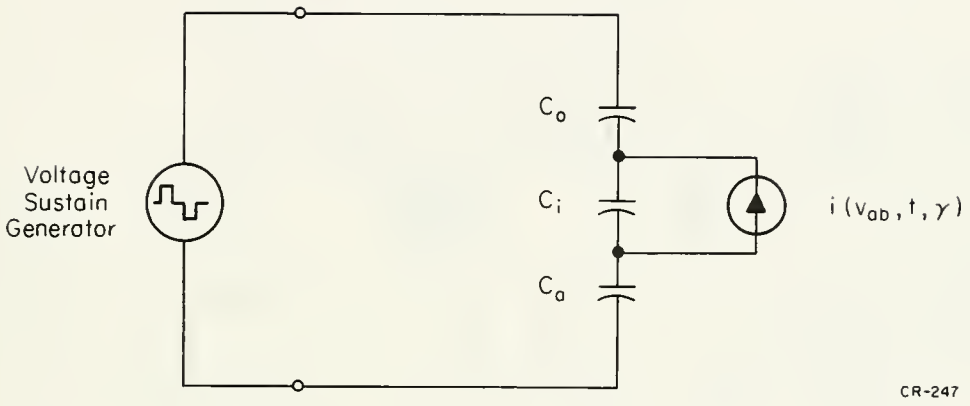


Fig. 9.7. Simplified model of a single plasma display cell.

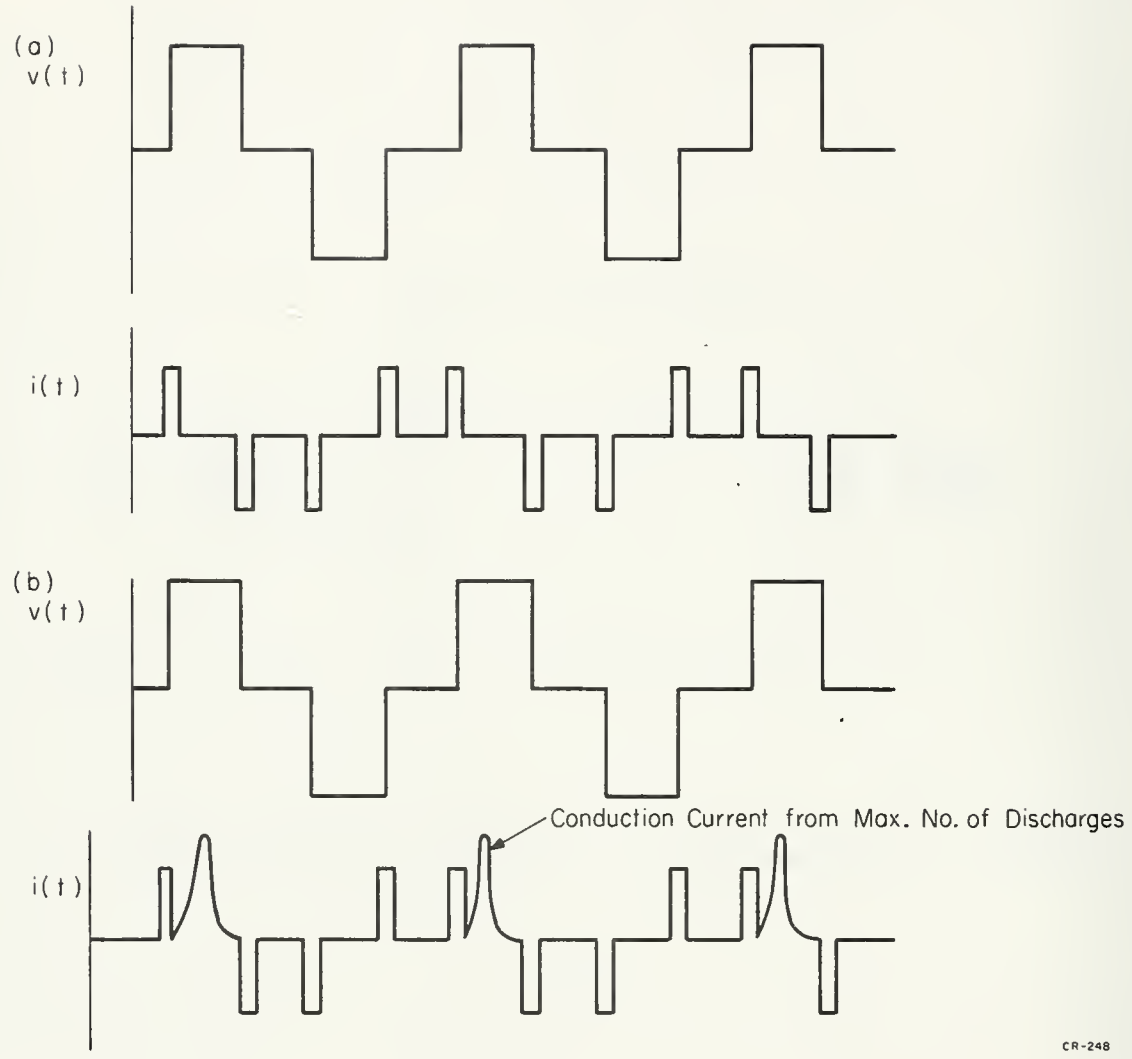


Fig. 9.8. Voltage and current waveforms provided by a plasma display panel generator.

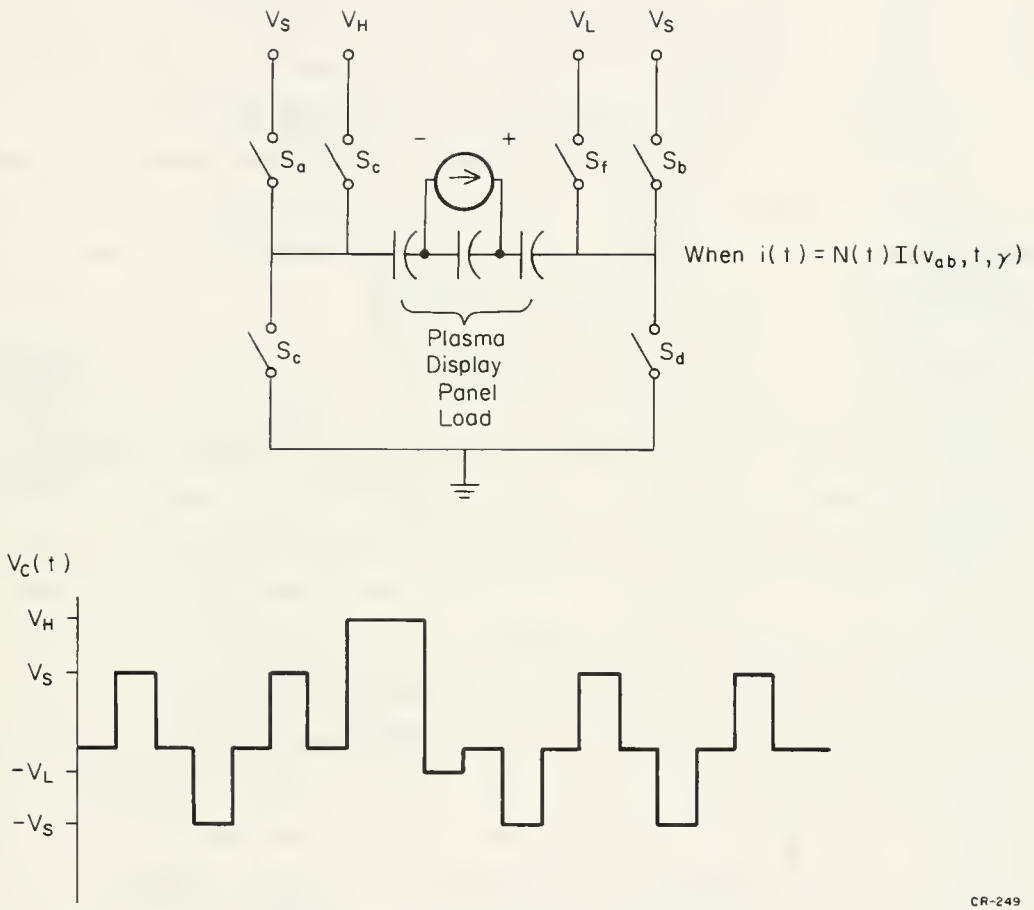


Fig. 9.9. An idealized model for a sustain voltage generator operating in the two-on-state mode.

the two-on-state, sustain-voltage waveforms described earlier.

Due to the reduced voltage requirements of the new prototype panels manufactured by Owens-Illinois, previously described address techniques are now being implemented through integrated-circuit, line-driver components such as those presently being used to drive conventional gas-tube, alpha-numeric displays. These line-driver schemes are then coupled with the sustain generators described earlier to form a complete plasma-display drive unit.

One method of operating a Plasma Display Panel employs transformers with multiple secondary windings and addressable voltage sources. The transformer secondaries are connected between the voltage sources and the panel electrodes so that the electrode voltage is the algebraic sum of two voltages. Each secondary on a particular transformer is in series with a different voltage source, so that if only one transformer primary and one voltage source are enabled, then a unique voltage is produced on a single electrode. The transformer primaries are driven only when writing or erasing. The voltage sources are the only source of sustaining voltage and are also used in addressing. This circuit was described in the CSL Progress Report for March-August 1968, and an abbreviated schematic is reproduced in Fig. 9.10 for reference.

It is possible to implement this system using any number of transformers, if it be assumed that the product of the number of transformers times the secondaries per transformer is the same as the number of electrodes being addressed. Since the number of voltage sources required is equal to the number of secondaries per transformer, the apparent number of output driving sources required to address n electrodes is $2\sqrt{n}$. However, the transformer primaries may be driven by $2\sqrt{N}$ sources, where N represents the number of transformers being used. Factors, such as the load on individual sustaining sources or transformers, and the relative cost of the different types of circuits, help determine how many transformers should be used. The chart in Table 9.1 shows some of the possibilities.

Circuits of this type have been constructed and are being used to operate a 100-mm panel (128X128 addressable cells). Evaluation so far indicates that erasing is a dependable process and is sufficiently tolerant of deviations in amplitude of both the sustaining signal and the erase pulse. For a particular Plasma Panel, Fig. 9.11 indicates the erasing characteristic of the entire system.

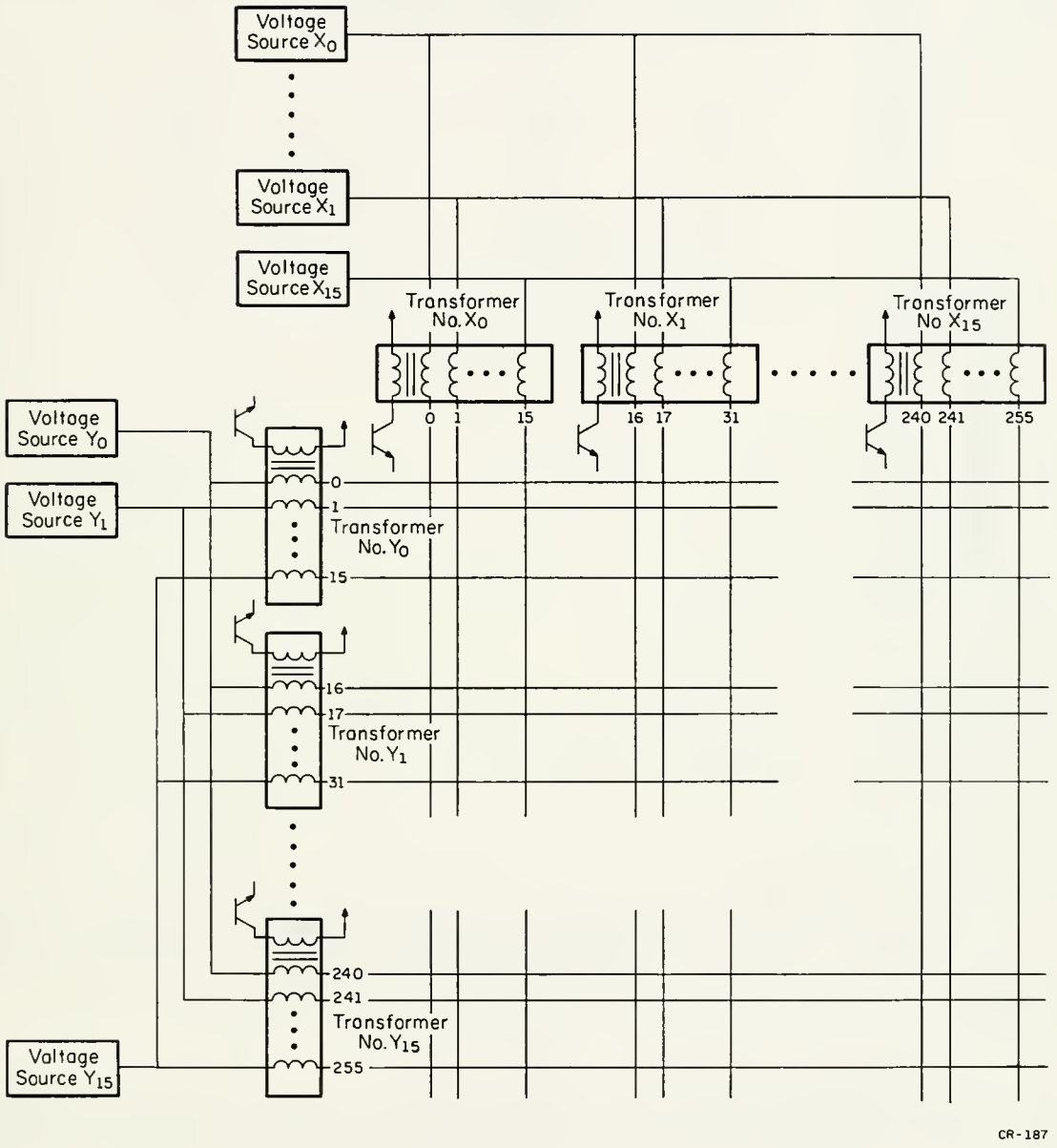


Fig. 9.10. Device-saving driving system for 256X256 Plasma Display Panel.

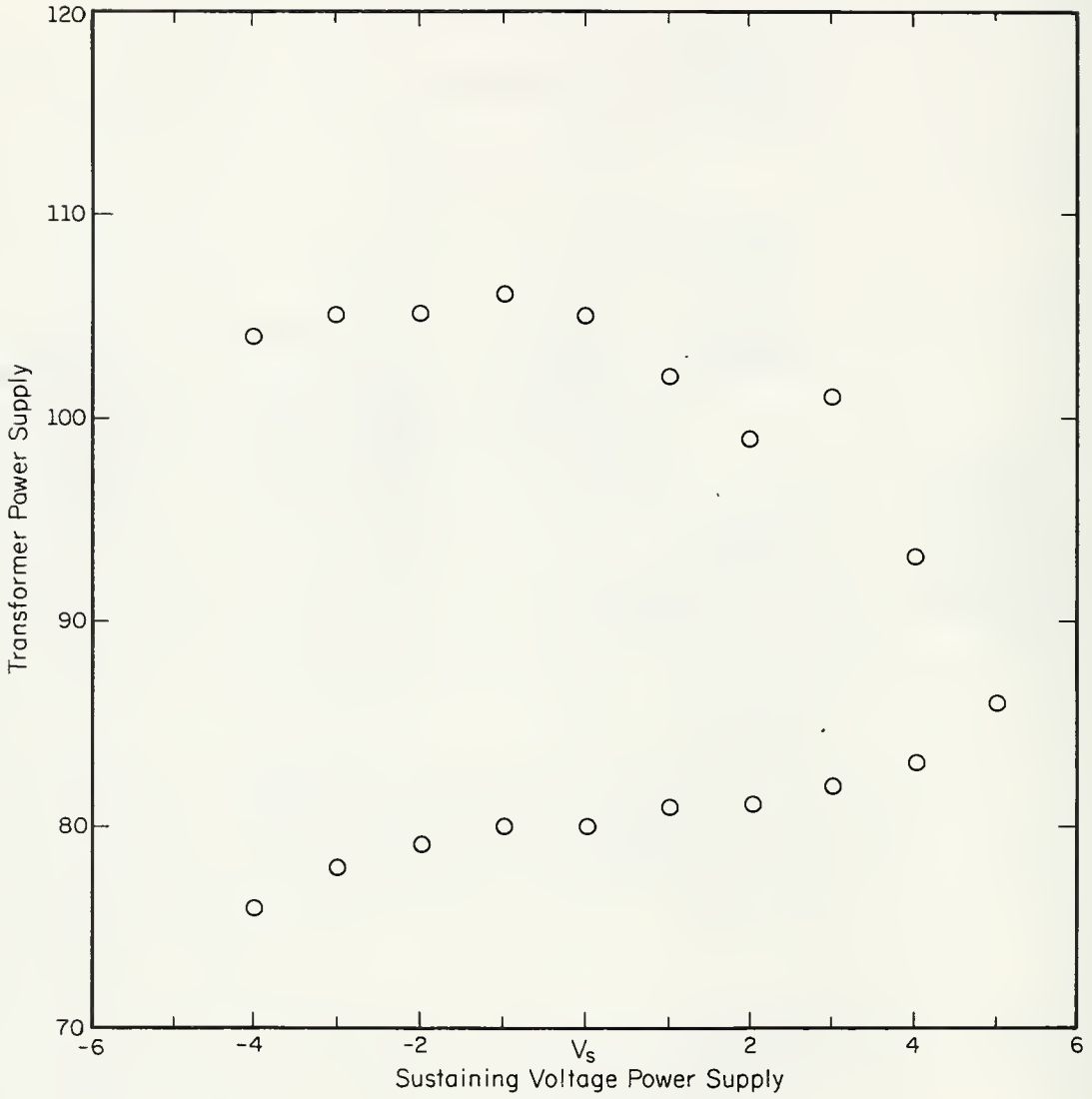


Fig. 9.11. Voltage range for successful erase with transformer driving system (Vs is expressed in volts deviation from center value).

Table 9.1. Transformer Connections

Number of Electrodes	Number of Transformers	Number of Secondaries per trans l	Number of Transformer Driving Sources	Number of Sustaining Drivers	Driving Sources for x or y	Total Driving Sources for Panel
128	64	2	8+8=16	2	18	36
128	32	4	8+4=12	4	16	32
128	16	8	4+4=8	8	16	32
512	64	8	8+8=16	8	24	48
1024	128	8	16+8=24	8	32	64

Writing is a more critical process. Some of the difficulty associated with writing is due to the fact that the probability of lighting a cell is considerably higher if other cells in the panel are already lighted. We assume that light from these cells produces the electrons, by photoemission, necessary to initiate a new discharge sequence. However, by having columns of spots lighted along edges of the panel, but outside the area being addressed, writing performance is satisfactory, though the margins are not as great as for erasing. A second version of this system now under construction employs the two-on-state mode (or a conditioning signal) and it is expected that this will greatly increase operating margins.

9.5. Field Studies

As technology on the Plasma Display becomes more advanced, and techniques utilizing the device become more sophisticated, a greater knowledge of the charge distributions and voltages inside the cell becomes desirable. To this end, a computer program is being developed to give the equipotential surfaces of the three-dimensional problem for the static case. Several problems with the more-or-less standard geometries have been solved and are presented.

At this stage of development all symmetries have been used. All cells are assumed to be in the same state and part of an infinite matrix of cells--hence the problem is periodic in x and y. Moreover, each cell is assumed to have the symmetry $\varphi(x,y,z) = -\varphi(y,x,-z) = \varphi(-x,y,z) = \varphi(x,-y,z)$ about its center which allows computation of only one corner (1/8th of a cell, see fig. 9.12). In future work some of these symmetries must be broken. In the z direction, which is not periodic, the potential on a plane at a distance Z from the conductor is approximated by a Fourier expansion subject to Laplace's equation.

Potentials are computed using the relaxation method, and the solution proceeds in two steps. A rough solution is generated using 10x10x30 matrix which can be stored in the computer core. A linear expansion is then done to a 40x40x120 (typical) or larger matrix, of which only three planes are stored in the computer core at one time, since there is not space for the whole array. The program then relaxes the center plane in core, writes the bottom plane on magnetic tape, and reads a new plane from another tape. Four tapes and special driver routines are used so that no time is lost in rewinding and input-output

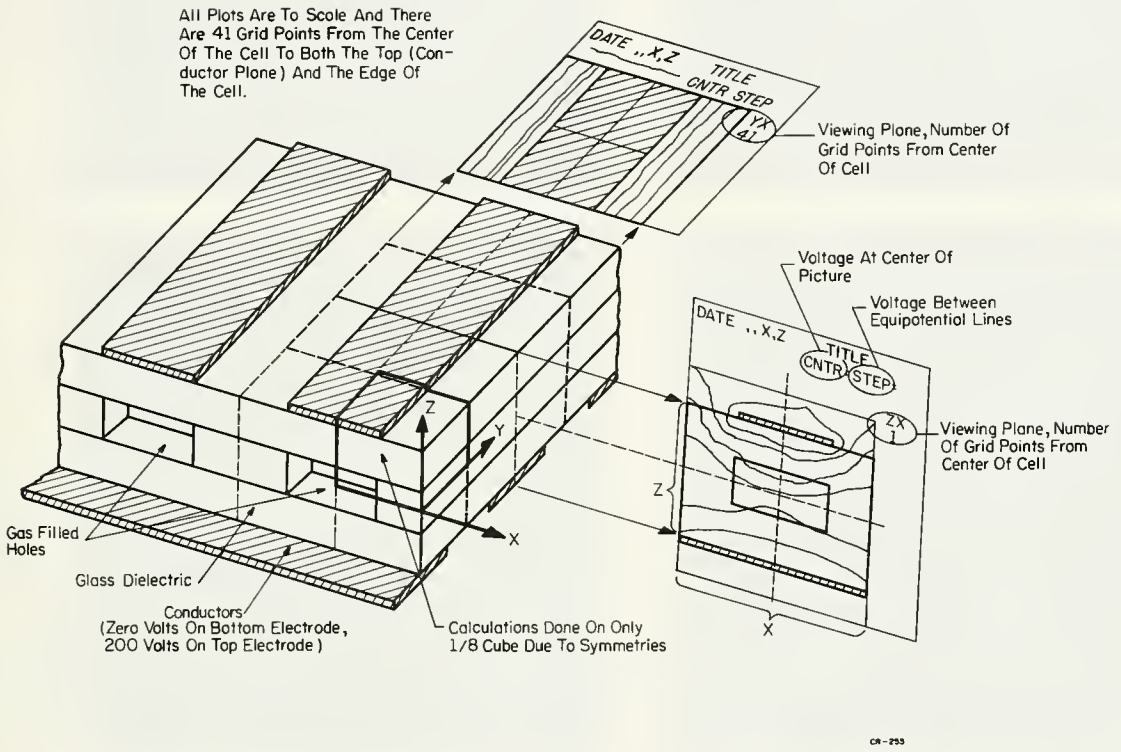


Fig. 9.12. Key to equipotential plots.

time is small compared to computation time. These methods were dictated by the computer facilities available. Since a typical problem takes considerable time, programs are being rewritten for more speed and flexibility.

Once a solution is generated, fields can be calculated and printed out with potentials in desired locations. To more effectively display solutions, a program was written to trace out the equipotential lines in a plane. Thus one can "slice" a cell and look at the equipotential lines resulting from the intersection of the slice and the equipotential surfaces within the cell. This is illustrated in Fig. 9.12. A sample view of a geometry which has been used in an actual cell at CSL is shown in Fig. 9.13.

An integral view of the relaxation process lies in noticing the averaging effect of Gauss's Law. That is the potential at the center is φ_c, which is

$$\bar{\varphi} = \int_s \varphi da / \int_s da \ , \qquad (1)$$

for a sphere.* When charges are present in the volume, Eq. (1) can be modified to

$$\bar{\varphi} = \varphi_c + \varphi_q \ , \qquad (2)$$

where $\bar{\varphi}$ is calculated as in Eq. (1). An approximation to Eq. (1) can be envisioned as replacing $\bar{\varphi}$ with the average of the potentials at the center of the faces of a cube, tangent to and surrounding the sphere.† In free space, the value of any point in a three-dimensional matrix thus becomes the average of the six points surrounding it. If this average is different from φ_c it is equivalent to having a charge in the region, as would be given by Gauss's Law. Now, what one might do physically is add an opposite charge at this point to remove this unwanted charge. However, this changes potentials over the entire problem, and its local effect depends in a complicated way on how the charge is distributed within the volume. Equation (2) says it is possible to add some charge distribution, the result summed up in the term φ_q, so that Eq. (1) is satisfied locally, and the rest of the problem is unchanged. Thus, what we do is form $\varphi_c^{new} = \varphi_q + \varphi_c$. The solution is still not correct over the entire matrix, since the same process applied at neighboring points will affect the result. However,

---

*see §9.6 for discussion.

†These ideas can be made rigorous. The approximation is good if the potential is essentially constant over the surface of the cube.

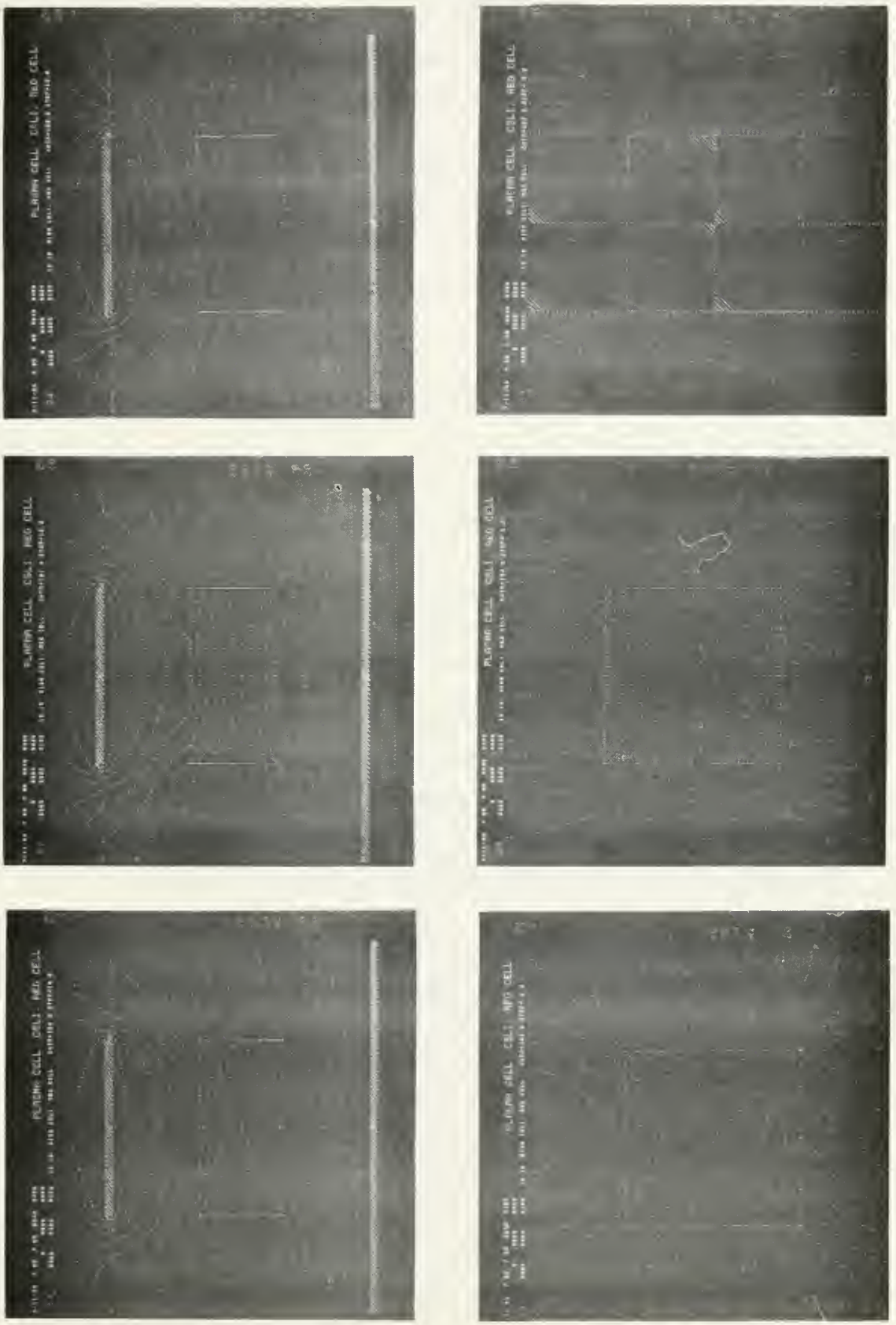


Fig. 9.13. Selected equipotential plots for a typical geometry.

after enough iterations over the entire matrix, the solution can be arbitrarily close. This process is equivalent to solving a set of simultaneous linear equations by an iterative method. Convergence for eliptic difference equations is guaranteed.

Convergence can be speeded up by adding $\beta\varphi_q$ (where β is called the over-relaxation factor, see Fig. 9.14) instead of just φ_q. In equation form, we thus have in free space

$$\varphi_c^{new} = \beta(\bar{\varphi}-\varphi_c)+\varphi_c \ ,$$

or

$$\varphi_c^{new} = \beta\bar{\varphi}+(1-\beta)\varphi_c \ ,$$

where

$$\bar{\varphi} = \sum_{i=1}^{3} [\hat{\varphi}_i]/6 \ ,$$

with $\hat{\varphi}_i=\varphi(x_i+h)+\varphi(x_i-h)$, for i=1,2,3 corresponding to x,y,z, and for h being the distance between grid points. If h is not the same in all dimensions, one can extrapolate to a cube using a quadratic. The result is

$$\bar{\varphi} = \tfrac{1}{2}(\hat{\varphi}_x+r_y^2\hat{\varphi}_y+r_z^2\hat{\varphi}_z)/(1+r_y^2+r_z^2) \ , \qquad (3)$$

with $r_y=h_x/h_y$ and $r_z=h_x/h_z$. This form is suitable for the computer.

It is interesting to point out that with the above approximation, Eq. (1) is equivalent to Laplace's equation for finite differences. That is, by rearranging terms in Eq. (1) with Eq. (3) we can obtain

$$D^2\varphi = \sum_{i=1}^{3} (\hat{\varphi}_i-2\varphi_c)/h_i^2 = 0$$

In the same manner, Eq. (2) is equivalent to Poisson's equation.

When charge distributions are present, the process is the same except that the charge term in Eq. (2) must be subtracted from $\bar{\varphi}$ to obtain the desired value of φ_c. That is, we must solve Poisson's equation instead of Laplace's. The exact form for this term is most easily derived from Poisson's equation, $\nabla^2\varphi=-\rho/\epsilon$. In the case of a three-dimensional charge distribution, we have immediately the finite difference equation $D^2\varphi=-\rho/\epsilon$. For a surface charge, say in the xy plane (z=z'), Poisson's equation can be written with a delta function $\nabla^2\varphi=-(\sigma/\epsilon)\delta(z-z')$(the Dirac delta has dimensions of reciprocal length), which

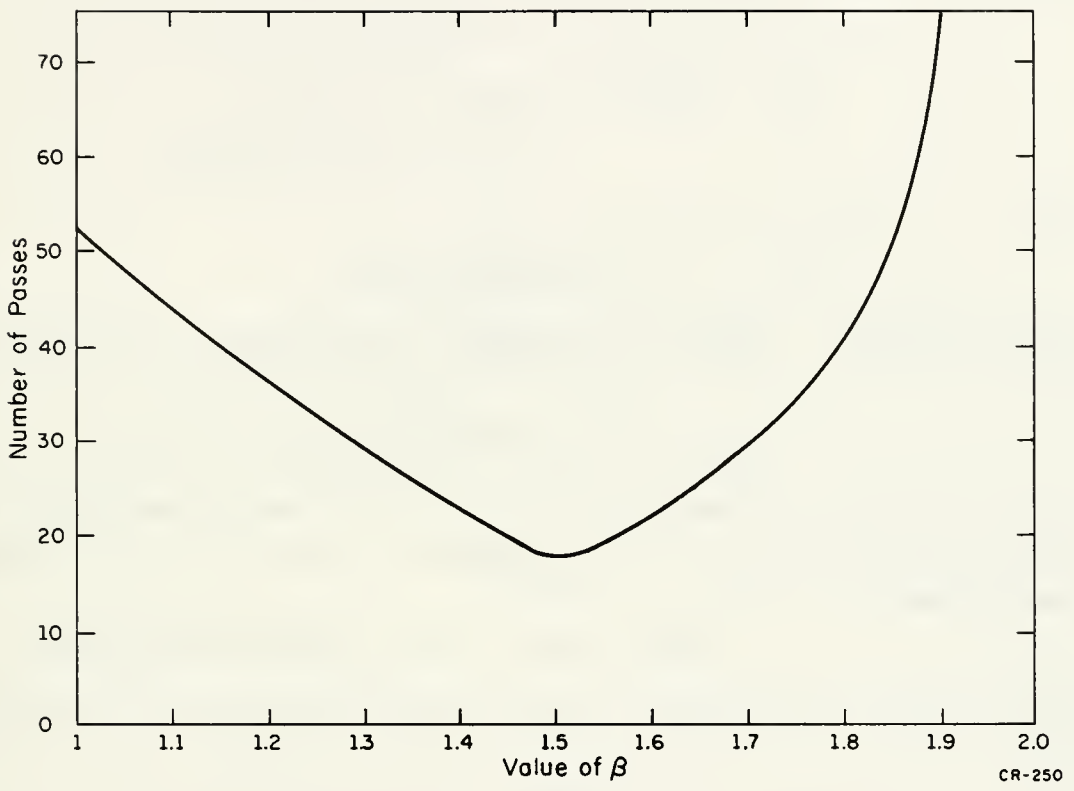


Fig. 9.14. The effect of a constant overrelaxation factor in three dimensions.

in its finite form becomes $D^2\varphi = (\sigma/\epsilon)\delta_{z,z'}/h_z$, where $\delta_{z,z'}$ is the Kronecker delta function.

For the case of induced charge at the boundary of dielectrics (refer to Fig. 9.15), the boundary condition is $E_2 = KE_1$, with $K = \epsilon/\epsilon_o$ being the dielectric constant, and the boundary charge is $\sigma/\epsilon_o = E_2 - E_1$, that is

$$\sigma/\epsilon_o = (1-1/K)E_2 = -(1-1/K)[\varphi(z+h)-\varphi_c]/h_z ,$$

$$\sigma/\epsilon_o = (K-1)E_2 = -(K-1)[\varphi_c-\varphi(z-h)]/h_z$$

$$= -(K-1)[\Delta\varphi_z-\varphi(z+h)+\varphi_c]/h_z ,$$

where

$$\Delta\varphi_z = \varphi(z+h)-\varphi(z-h) .$$

Combining the two, we obtain a form which is stable during the iteration process: with

$$h_z\sigma/\epsilon_o = A = (1-K)/(1+K) = A\Delta\varphi_z .$$

It was found that the solution at the boundries converged more slowly than in the other regions. A function was therefore invented to accelerate convergence at the boundaries. We let

$$h_z\sigma/\epsilon = A'\Delta\varphi_z \qquad A' = a+(1-|A|)(k-K)/(k+K)$$

where k is the "instantaneous" value of K (ratio of fields) during the calculation.

We thus obtain for the various charge distributions the following form:

$$\varphi_c^{new} = \tfrac{1}{2}\beta(\hat{\varphi}_x+r_y^2\hat{\varphi}_y+r_z^2\hat{\varphi}_z-\Sigma)/(1+r_y^2+r_z^2) + (1-\beta)\varphi_c ,$$

where

$\Sigma = \rho h^2/\epsilon$ for spatial distributions, $\Sigma = r_z^2 h\sigma\delta_{z,z'}/\epsilon$ for z-plane distributions, and $\Sigma = r_z^2 A'\delta_{z,z'}\Delta\varphi_z$ at dielectric boundaries in z.

It is worth mentioning that at this time we have considered only rectangular dielectric boundaries. A more general form of the above has been derived for arbitrary surfaces in three dimensions and programs utilizing it are being developed. This will be presented along with results at a later date.

9.6. Discussion of Eq. (1)

In Eq. (1), the symbol s denotes the sphere as the range of the surface integral; this equation may be written

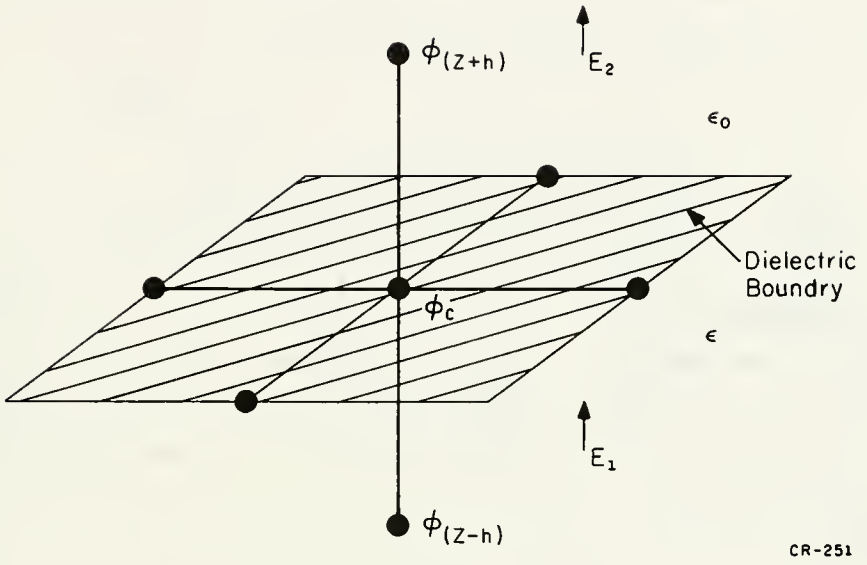


Fig. 9.15. Relaxation over a dielectric boundary.

$$\bar{\varphi} = (1/4\pi r^2)\int_s \varphi da = (1/4\pi)\int_s \varphi d\Omega \ ,$$

where the differential solid angle $d\Omega = da/r^2$ is introduced to indicate the nature of the r dependence. Now we differentiate with respect to r and obtain

$$d\bar{\varphi}/dr = (1/4\pi)\int_s (d\varphi/dr))d\Omega = -(1/4\pi r^2)\int_s E_r da \ .$$

The integral is zero in free space by Gauss's Law and $\bar{\varphi}$ is therefore independent or r. Now as $r \to 0$, $\varphi \to \varphi_c$, and hence we have $\bar{\varphi}=\varphi_c$.

If this were not free space, the result would be

$$\bar{\varphi} = \varphi_c + \int_r [\int_v \rho dV] dr/4\pi r^2 = \varphi_c + \varphi_q \ ,$$

where φ_q depends on the charge distribution and the radius.

9.7. Stability Theory

In an earlier report, we discussed the conditions for stability of discharge times in a Plasma Display cell in terms of the relation between change in wall voltage after a discharge and the slope of the exciting voltage at the time of the discharge. When the exciting voltage across a cell changes rapidly enough, however, the discharge will not mature until after the voltage reaches its peak value. It is then more convenient to discuss the stability directly in terms of the influence of cell voltage on the transfer of charge. In either case, of course, the principles are the same. The charge transferred increases with the cell voltage, or more precisely, with the extent to which the cell voltage exceeds the breakdown voltage. In the first case, this overvoltage is controlled by the slope of the exciting voltage at the time of firing. In the second case, it is controlled directly by the magnitude of the exciting voltage and the wall voltage at the time of firing.

At the i^{-th} discharge in a sequence, the total change in wall voltage, $2V_{ti}$, is a function of the cell voltage V_{ci}, which is, in turn, the sum of the exciting voltage V_s and the wall voltage V_{wi}, as follows:

$$2V_{ti} = f(V_{ci}) \ , \tag{1}$$

$$V_{ci} = V_s + V_{wi} \ . \tag{2}$$

Figure 9.16 shows the relations among these voltages for a square-wave sustaining signal. The relation between two successive values of wall voltage is

$$V_{wi+1} = -V_{wi} + 2V_{ti} \ , \tag{3}$$

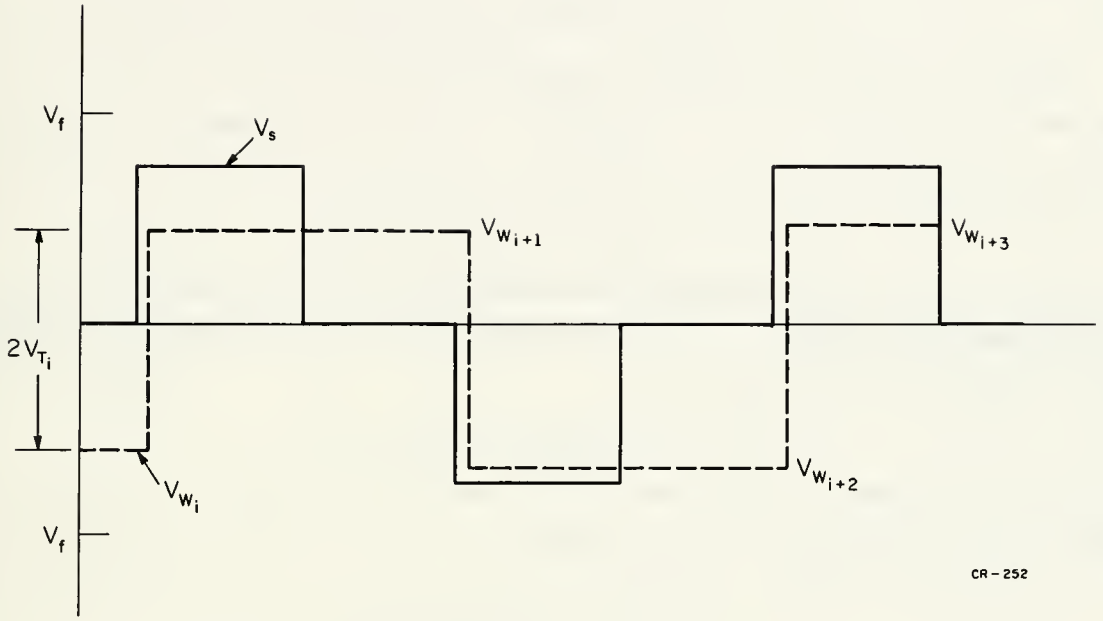


Fig. 9.16. Approach to equilibrium in plasma display cell.

or with Eq. (1)

$$V_{wi+1} = -V_{wi}+f(V_{ci}) \quad . \tag{4}$$

If the process is stable, the wall voltage V_{wi} and the cell voltage V_{ci} approach their equilibrium values,

$$\begin{aligned} V_{wi} &= V_{ci+1} = V_{wo} \quad , \\ V_{ci} &= V_{ci+1} = V_{co} \quad , \end{aligned} \tag{5}$$

and the total change in wall voltage at a discharge is twice the magnitude of the wall voltage

$$2V_{ti} = 2V_{to} = 2V_{wo} = f(V_{co}) \quad . \tag{6}$$

In terms of the difference δ_i between the actual wall voltage V_{wi} and the equilibrium wall voltage V_{wo}, and the difference δ_{i+1} between V_{wi+1} and V_{wo}, this relation may be written

$$V_{wo}+\delta_{i+1} = -(V_{wo}+\delta_i)+f(V_{co}+\delta V_c) \quad . \tag{7}$$

We note that since the magnitude of the sustaining voltage is constant, the difference between V_{ci} and V_{co} is also δ_i.

To first order, the change in wall voltage at the i^{-th} discharge is

$$2V_{ti} = f(V_{co})+(\partial f/\partial V_c)\delta_i \quad , \tag{8}$$

which may be inserted in Eq. (7), which then is

$$V_{wo}+\delta_{i+1} = -(V_{wo}+\delta_i)+f(V_{co})+(\partial f/\partial V_c)\delta_i \tag{9}$$

or after rearrangement of terms,

$$\delta_{i+1} = (-1+\partial f/\partial V_c)\delta_i+f(V_{co})-2V_{wo} \quad . \tag{10}$$

Since the sum of the last two terms is zero from Eq. (6) we finally obtain $\delta_{i+1}=(-1+\partial f/\partial V_c)\delta_i$, or with $\alpha=\partial f/\partial V_c$, we write this as

$$\delta_{i+1} = (-1+\alpha)\delta_i \quad . \tag{11}$$

For values of α in the range

$$0 < \alpha < 2 \tag{12}$$

the perturbations in wall voltage V_{wi} (and in cell voltage V_{ci}) damp out, and the process is stable.

The damping of a perturbation increases as α approaches the value $\alpha=1$, and will, in fact, be completely corrected on the next successive discharge when $\alpha=1$. For most display applications, therefore, this is the best value for α. When $\alpha=2$ at the edge of stability, a perturbation once introduced persists

indefinitely. This case is illustrated in Fig. 9.17 in which V_{wa} and V_{wb} represent two stable states. The other limiting case, in which $\alpha=0$, implies that a change in wall voltage is independent of cell voltage, and is not a practical condition for a Plasma Display cell.

If the applied voltage V_s is increased, and if the process is stable, the wall voltage will approach a new, and larger, equilibrium value. Since the cell voltage is the sum of the applied voltage and the wall voltage,

$$V_c = V_s+V_w \ , \tag{13}$$

the increase in V_c must exceed the increase in V_s. The process, then, that relates V_w to V_s is, even if stable, a regenerative one.

We assume that, corresponding to an applied voltage, V_{a1} is a wall voltage V_{w1}, and similarly V_{w2} corresponds to V_{a2}. Then Eq. (1) for each of the two cases is

$$2V_{t1} = f(V_{c1}) \ , \tag{14}$$

and we have

$$2V_{t2} = f(V_{c2}) \ . \tag{15}$$

With

$$V_{w2}-V_{w1} = \Delta V_w$$

and (16)

$$V_{s2}-V_{s1} = \Delta V_s \ ,$$

Equation (14) can be written in the form

$$\begin{aligned} 2V_{t1}+2\Delta V_w &= f(V_{c1}+\Delta V_w+\Delta V_a) \\ &= f(V_{c1})+(\partial f/\partial V_c)\Delta V_c \ , \end{aligned} \tag{17}$$

to first order. Since $\Delta V_c=\Delta V_s+\Delta V_w$, we obtain

$$2V_{t1}+2\Delta V_w = f(V_{c1})+(\partial f/\partial V_c)\Delta V_s+(\partial f/\partial V_c)\Delta V_w \ , \tag{18}$$

and since $2V_{t1}=f(V_{c1})$, then

$$2\Delta V_w = (\partial f/\partial V_c)\Delta V_s+(\partial f/\partial V_c)\Delta V_w \ , \tag{19}$$

and with some rearrangement of terms

$$\frac{\Delta V_w}{\Delta V_s} = \frac{\partial f/\partial V_c}{2-\partial f/\partial V_c} \ . \tag{20}$$

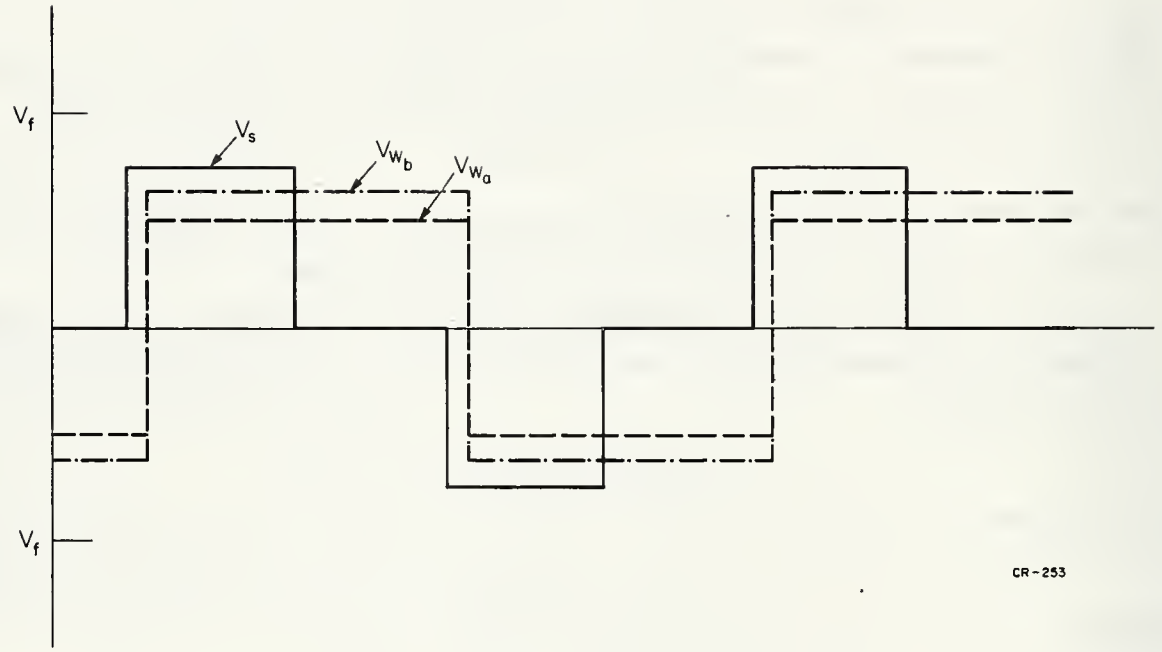


Fig. 9.17. Two persistent states with $\alpha = 2$.

With

$$\alpha = \partial f/\partial V_c ,$$

we write

$$\frac{\Delta V_w}{\Delta V_s} = \alpha/(2-\alpha) . \qquad (21)$$

Again we see that $\alpha=2$ defines the boundary of stability. For the ideal case for which $\alpha=1$, $\Delta V_w/\Delta V_s=1$. Figure 9.18 illustrates this case. The ratio $\Delta V_w/\Delta V_s$ is an important parameter for a Plasma Display Panel, operating in this mode, and fortunately it is easily measured.

9.8. Phosphors in the Plasma Display

The addition of phosphors to the Plasma Display Panel could, in principle, enhance the brightness of the Plasma Panel, and it could make possible the fabrication of a multicolor display. Stredde[1], in his thesis investigation, found, in fact, that with some phosphors in a xenon discharge the brightness was, in fact, greater than that observed in a neon nitrogen discharge without a phosphor. No quantitative information was obtained at that time.

In preliminary work during the past year, we observed that those phosphors designated as ultraviolet were brighter when placed in a Plasma Display discharge than those phosphors designated as cathodoluminescent. This result is consistent with the result of two other experiments that were performed recently. In one, a phosphor layer was excited in a vacuum by a beam of electrons. Until the accelerating voltage was raised to about 1500V, considerably in excess of the electron energies to be found in the Plasma Display Panel, the phosphors showed little response. In the second experiment the phosphors were separated from the Plasma Display discharge by a quartz panel. The resulting luminesence, due entirely to ultraviolet radiation, was acceptable for display purposes.

Several fabrication problems encountered in the early stages of this work have inhibited progress. The first panels were made with two center structures each perforated with holes. The holes of one of these panels were stuffed with phosphor, and the holes of the second panel, placed adjacent to the first, defined the discharge space. When these structures were evacuated, they would often crack. Furthermore, the nonuniform thickness of the phosphor layers led to differences in brightness. The phosphors are now being settled from a

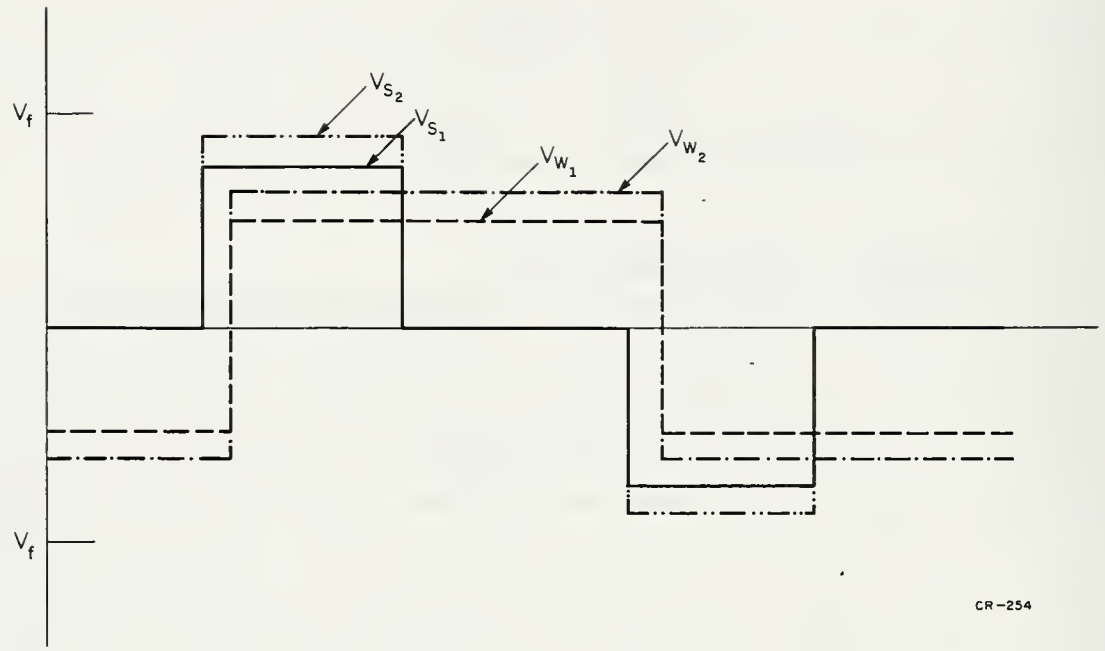


Fig. 9.18. Equilibrium states for two values of Vs with $\alpha = 1$.

colloidal suspension, much as they are in the television industry. Furthermore the experimental panels are now first assembled in a frame and then the entire structure is placed in a vacuum chamber before it is evacuated and filled with gas. In this way, there is no pressure differential between the inside and the outside of the panel at any time.

Another, rather serious, problem became apparent with work on some of the oxide phosphors. After exposure to the Plasma Display discharge, the efficiency of these phosphors deteriorated to a point where they were no longer useful in a display. We are now working primarily with rare-earth phosphors provided by the General Telephone and Electronics Company, and we have not observed any deterioration up to 98 hours thus far.

The essential problems in developing a phosphor Plasma Display Panel are to optimize the quantum efficiency of the gas-phosphor combination, and to minimize losses of radiation within the phosphor layer itself. In experiments now being planned, the influence of these parameters on the brightness and on the memory properties of the Plasma Display Panel itself will be studied.

9.9. References

1. Stredde, E., The Development of a Multicolor Plasma Display Panel, CSL R-Report, R-370, November, 1967.

J. Voss
J. Akin
J. Bouknight
W. Brown
D. Erbes*
D. Hansen*
J. Hansen*
M. Jones*
L. Kadanoff
M. Miller
M. Nicholas
B. Payne*
M. Plunk*
M. Vaughn*

10.1. Introduction

This project is a study of urban growth patterns as they are manifested in changing land use and property value. The plan is to obtain a rich and full data base about a particular town during its entire history--including prices for the transfer of real estate of half blocks throughout the urbanized area--and then use this data to find out as fully as possible the factors which shaped this observed pattern of growth and development. Kankakee is the city we have chosen for this study.

Notice that in some respects our procedure is different from the one most commonly employed in quantitative social scientific investigations. Often the investigator starts from a hypothesis: for example, that house prices tend to fluctuate in racially changing neighborhoods and then proceeds to design a method of obtaining a carefully controlled set of data so as to test this hypothesis. We are working in the opposite direction. We seek a broad data base which may be surveyed so that a wide variety of specific hypotheses may be formed and tested numerically.

10.2. Collection of Data on Price

At the beginning of the report period, we had collected from the files of the Kankakee County Title and Trust Company data of each sale of real estate on half blocks in the urbanized area of Kankakee (see Fig 10.1 and 10.2). These data include subdivision, block and lot number transacted, date of recording, recorded price, type of deed, and the book and page number of the record. These data describe roughly 30,000 transactions.

---

†Supported by the Ford Motor Company.

*Undergraduate Students

Fig. 10.1. State of Illinois. Circle indicates location of Kankakee. Source: U.S. Department of Interior, Geological Survey.

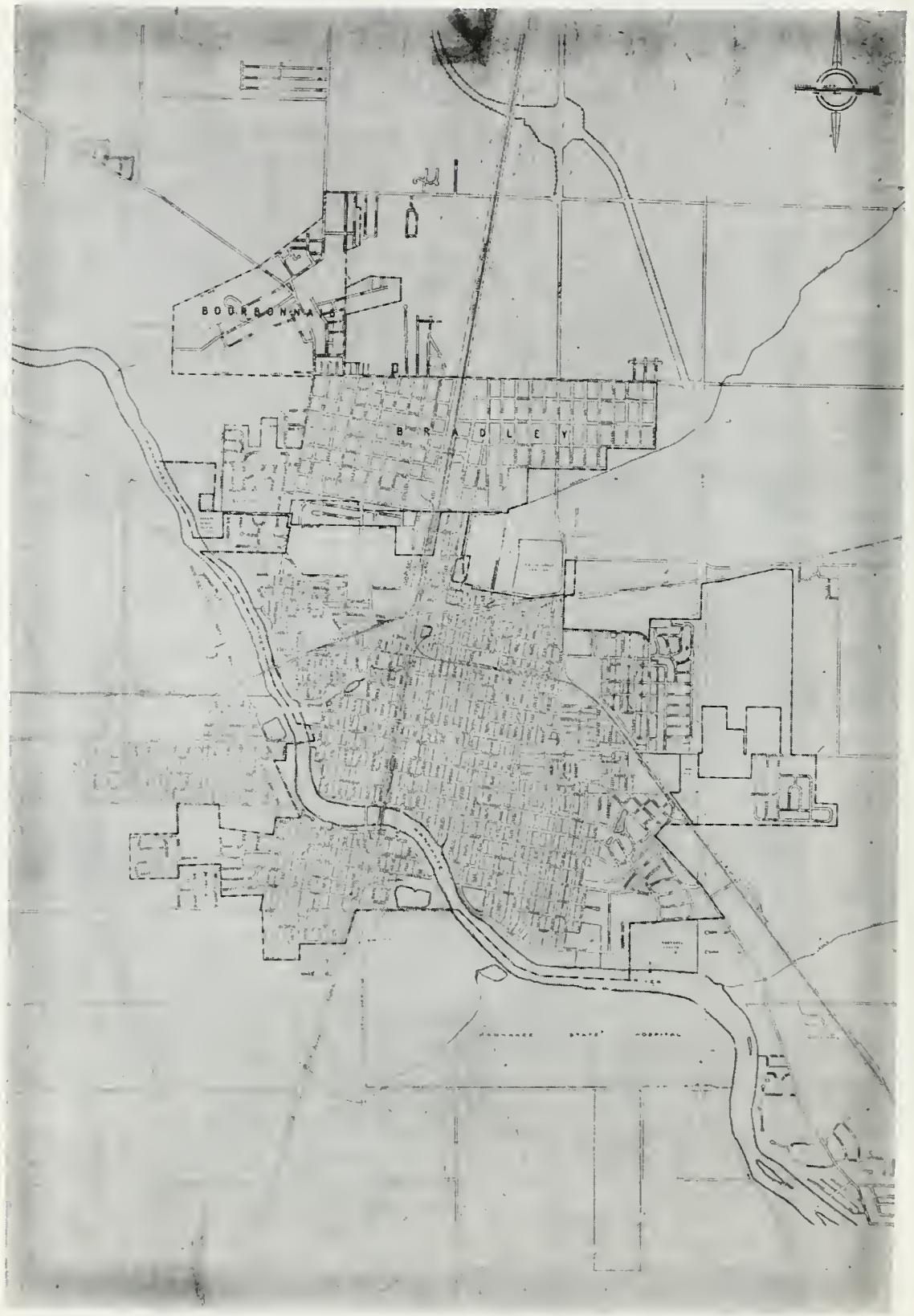


Fig. 10.2. Kankakee, Illinois. Source: Harlan, Bartholomew and Associates, St. Louis, Missouri.

We added to this information a computed area of transacted land and codes to indicate the streets abutting the land as well as the parcels' positions within the block.

All these data were converted into punched cards and read onto magnetic tape.

Many of the transactions indicated a dummy price: $1, $10, or nothing. To obtain the actual price for these transactions, we looked up microfilm copies of the original deeds in the Kankakee County Courthouse. Our work here was made possible by the enthusiastic cooperation of the past and present county recorders: H. K. Nally and B. Weiner. The original deeds have affixed to them federal or state tax stamps at a rate of $1 per thousand dollars of original transaction. By this means, we were able to ascertain the price for some 6,000 transactions which had an unrealistic price listed in the abstract office.

The collection of price data is now completed. The data are available in a "matrix printout" (Figs. 10.3 and 10.4) which lists each transaction in a given block. A price is printed opposite the date of transaction and below each of the lots involved in the sale. The symbols M, T, W, and D indicate respectively a Mortgage, Trust Deed, Warantee Deed, and a Deed. The symbol S indicates data taken from tax stamps. The symbol P shows that only part of the lot has been sold.

10.3. Street and Block Data

Presently we have nearly completed the task of recording coordinates for each of the following:

a. subdivision boundaries
b. center lines defining alleys, streets, and highways
c. railroads
d. river
e. city boundaries
f. bridges.

For all items except d, dates of construction and destruction (or revision) are included. Data and dates for the size, shape, and orientation of each block have also been taken. These data have been taken from city maps and the plat records obtained from the County Recorder's office.

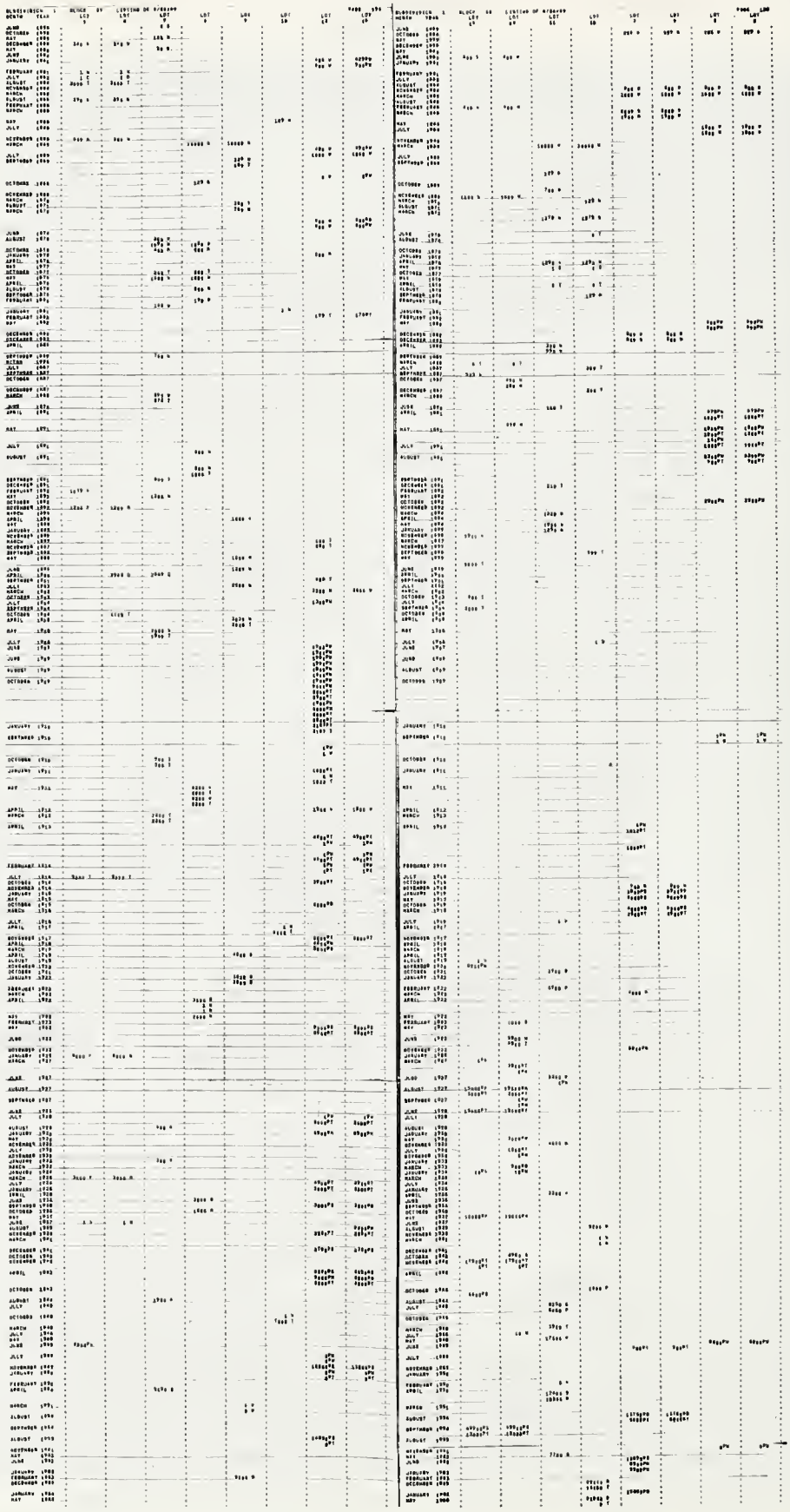


Fig. 10.3. A "matrix printout" listing all price data for block 29 of subdivision 1. The lines list transactions with dates at the left and center. The remaining 16 columns are for each lot in the block with price when sold. Data source: Kankakee Title and Trust Company and County Recorder's Office.

SUBDIVISION 1 BLOCK 26 LISTING OF 6/26/69 PAGE 19A

MONTH	YEAR	LOT 1	LOT 4	LOT 5	LOT 8	LOT 9	LOT 12	LOT 13	LOT 16
JUNE	1858			0 L					
OCTOBER	1858								
MAY	1859			100 W					
DECEMBER	1859	340 W	340 W						
MAY	1860			50 W					
JUNE	1860								
JANUARY	1861							425 W	425PW
								500 W	500PW
FEBRUARY	1861	1 W	1 W						
JULY	1862	1 C	1 O						
AUGUST	1864	3000 T	3000 T						
NOVEMBER	1864								
MARCH	1865								
AUGUST	1866	350 W	350 W						
FEBRUARY	1868								
MARCH	1868								
MAY	1868						125 W		
JULY	1868								
NOVEMBER	1868	600 W	800 W						
MARCH	1869				10000 W	10000 W			
								450 W	450PW
JULY	1869							1000 W	1000 W
SEPTMBER	1869					125 W			
						150 T			
OCTOBER	1869				125 W			0 W	0PW
NOVEMBER	1869								
MARCH	1870								
AUGUST	1871					350 T			
MARCH	1872					750 W			
								600 W	800PW
								800 W	800PW
JUNE	1874								
AUGUST	1874			300 W					
				1050 W	1050 W				
OCTOBER	1874			400 M	400 M				
JANUARY	1875							200 M	
APRIL	1876								
MAY	1877								
OCTOBER	1877			200 T	200 T				
MAY	1878			1200 W	1200 W				
APRIL	1879								
AUGUST	1879				500 W				
SEPTMBER	1879								
FEBRUARY	1880				150 W				
				125 W					
JANUARY	1881						1 W		
FEBRUARY	1882							175 T	175PT
MAY	1882								
DECEMBER	1882								
DECEMBER	1883								
APRIL	1885								
SEPTMBER	1885			700 W					
MARCH	1886								
JULY	1887								
SEPTMBER	1887								
OCTOBER	1887								
DECEMBER	1887								
MARCH	1889			850 W					
				875 T					
JUNE	1890								
APRIL	1891								
MAY	1891								
JULY	1891								
					400 W				
ALGUST	1891								
					500 W				
					1000 T				
SEPTMBER	1891			600 T					
DECEMBER	1891								
FEBRUARY	1892	1075 W							
MAY	1892			1300 W					
OCTOBER	1892								
NOVEMBER	1892	1200 W	1200 M						
MARCH	1894								
APRIL	1894					1000 W			
MAY	1894								
JANUARY	1895								
NOVEMBER	1895								
MARCH	1897							100 T	
NOVEMBER	1897							250 T	
SEPTMBER	1898								
MAY	1899					1200 M			

Fig. 10.4. Detail of Fig. 10.3. Each line lists the price for a given transaction whenever the lot indicated is involved.

These data have been only partially converted to machine-readable form, and only a limited amount of error elimination has been carried out. However, the work we have done so far has permitted us to construct street maps like those shown in Figs. 10.5 and 10.6 which provide a continually changing city map as a function of time.

This dynamic map is displayed on the CRT connected to the Laboratory computer. The dynamic mapping program developed during this reporting period permits data both to be added to and subtracted from the map as the city develops. This is a major advance over the "add-only" program developed earlier. Of course, the dynamic map will serve as a framework for computer-generated displays of price and land-use data within each block.

10.4. Land Use

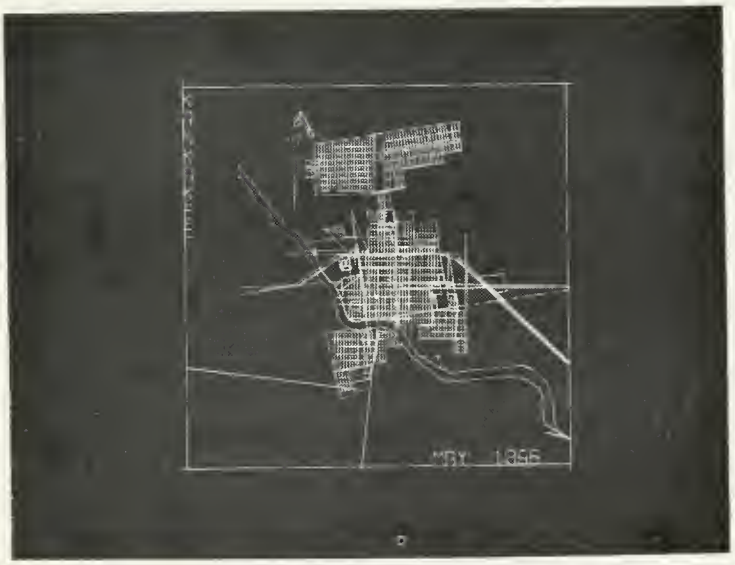
With the aid of Sanborn Insurance Maps, as well as other maps and aerial views, land use and structure data can be assembled for the following years: 1869, 1870, 1887, 1890, 1892, 1896, 1901, 1907, 1914, 1923, 1930, 1953, 1956, 1959, 1961, and 1965. These data then permit us to record the number of square feet within each block devoted to the following uses:

a. residential
b. commercial
c. industrial
d. public

as a function of time. The land use inventory should be completed in the fall of 1969.

10.5. Histories of Individual Dwellings

The use of the sequenced maps permit the development of the history of individual structures. From the maps and the internal evidence of price, we can infer when each dwelling was constructed and how its use has changed. Each dwelling on a block is assigned a number and record is made of its date of construction, the type of dwelling (two-family house, apartment house, flat, etc.), and its date of demolition. This task should also be completed in the fall of 1969.



Figs. 10.5, 10.6. Preliminary computer-generated map of Kankakee metropolitan area. Streets, railroads, subdivision boundaries and the river are shown at different times. These pictures are incomplete and contain several errors.

10.6. Restructuring of Price Data

Using the history of each residential structure and the structures included in the transaction, we can turn to the economic history of the market for vacant and residential land in Kankakee.

If the land sold is vacant, a notation to that effect will be made in the "land-use code" on the data record. If a residential structure or structures are included, this will instead be indicated as a residential land use. In the latter case, a notation will be made on the transaction record indicating the particular structures being sold. Figure 10.7 shows a matrix printout marked up to indicate these new data. Figure 10.8 shows a Sanborn Insurance Company map used to find these land uses as a function of time and to identify particular residential structures.

10.7. Transportation Data

The city of Kankakee has descriptive information on streets and street improvements dating back at least to the turn of the century, and we are in the process of assembling this information. In addition data are being collected on all public transit systems that have existed in Kankakee.

10.8. Educational Information

Data are being assembled about the history of every school in Kankakee, including its school district, enrollment, and a quality estimate.

10.9. Display and Analysis

During the next year, we shall devote most of our attention to the problems of analyzing the data we have collected. The analysis will focus on how the six basic variables

1. the price of homes
2. the rate of change of home price
3. the price of vacant land
4. the rate of change of the vacant-land price
5. the density of residential development
6. the rate of construction of new homes,

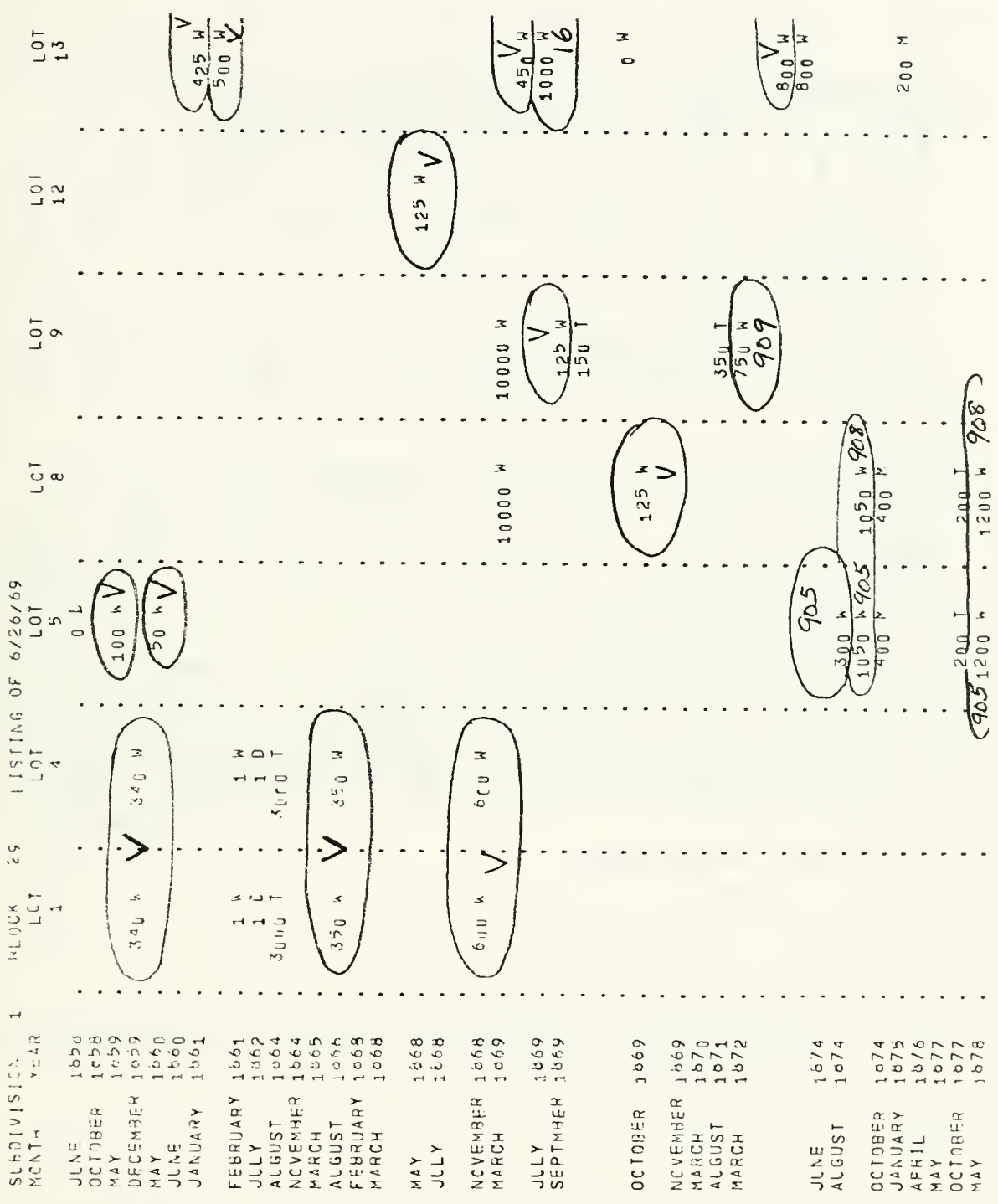


Fig. 10.7. "Matrix printout" marked to indicate the property transferred in the various transactions. The V indicates vacant land, and the written numbers denote particular residential structures within the block. The circles enclose valid transactions which transfer residential or vacant land.

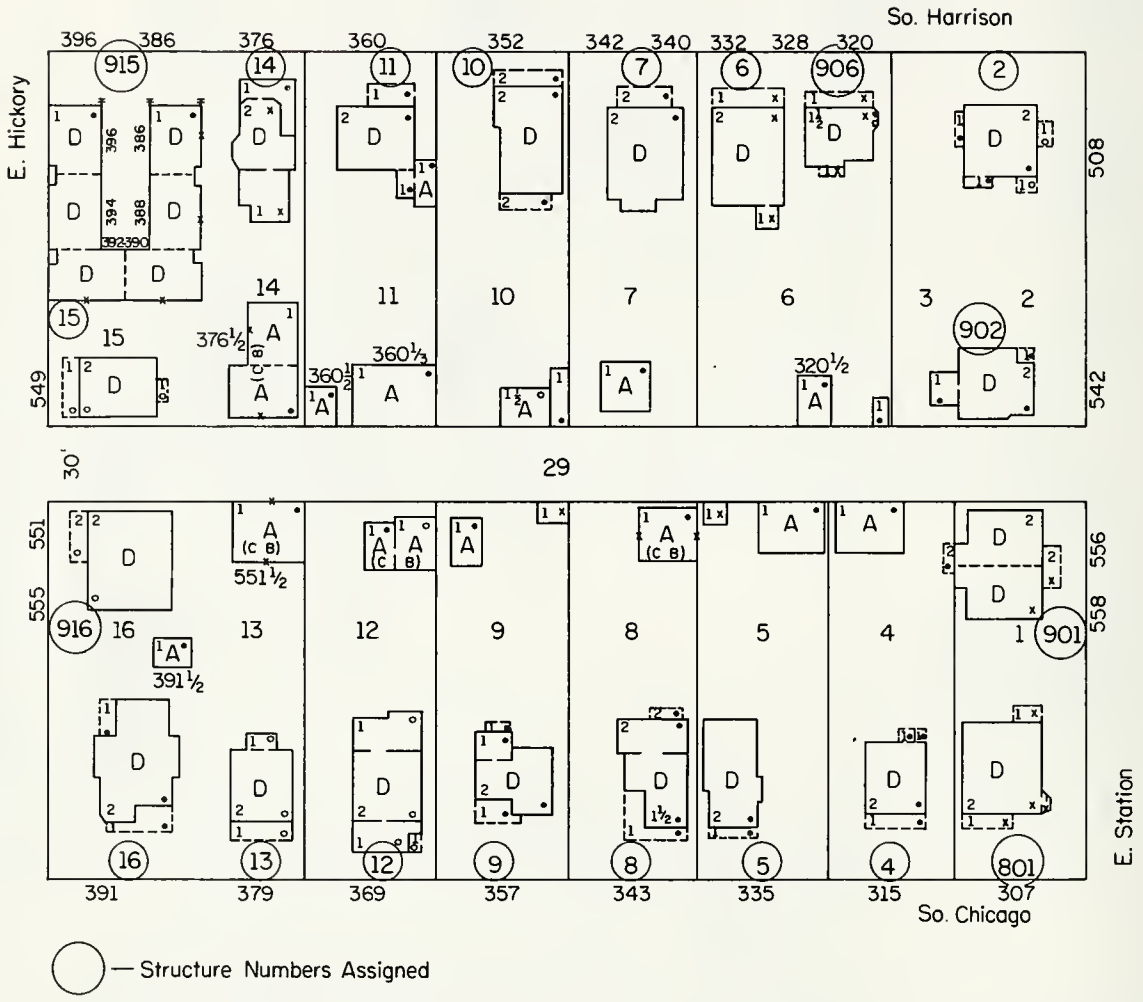


Fig. 10.8. Sanborn Insurance Company map of subdivision 1, block 29. The circled numbers indicate our numbering of residential structures within the block.

are affected by various forces in the neighborhood and community. Our data will enable us to study how variables 1-6 are affected by

a. time, over-the-road, and airline distance from the commercial and industrial centers
b. the quality of the schools
c. the proximity of scenic amenities (e.g., the river)
d. the values of variables 1-6 in the immediate neighborhood
e. mixture of industrial and commercial uses on the block
f. density of residential development on the block
g. diversity of home values on the block
h. ages of the homes on the block
i. availability of public transportation,as well as many other possibly causative variables.

The analysis of these relationships will be facilitated by the possibility of displaying spots of light for the various kinds of transactions in the various price ranges on the dynamic map of the town. The display will help us see the relationships among the variables in question. By using the display, we can visually scan the price developments and correlate these with the important variables. Then, a factor, or multiple-correlation, analysis can be used to pin down the relationships in a precise fashion.

R. T. Chien
R. B. Ash
L. Bahl
Y. Chang
D. Cooper
T. Cordaro
R. Cullum
P. Davis
A. Haddad
C. Hartmann
S. J. Hong
R. J. Kubala
E. Lyman
E. Mark
J. Moore
A. Moy
S. Ng
R. Pavek
F. P. Preparata
K. Tzeng

11.1. Introduction

In this section, we report our progress in the area of information-processing techniques and systems. Topics discussed include methods of achieving systems reliability through coding, design techniques of digital systems, estimation and filtering, source coding and time-continuous channels, information retrieval and communications.

R. T. Chien

11.2. Algebraic Coding Theory

11.2.1. Iterative Decoding of BCH Codes

This work is largely concerned with the decoding problem of error-correcting codes. A detailed account is given in the technical report R-404 of the Coordinated Science Laboratory.

A direct approach to the solution of Berlekamp's key equation for decoding Bose-Chaudhuri-Hocquenghem codes with the consequent derivation of a modified Berlekamp Iterative Algorithm is presented. Iterative decoding of a syndrome sequence in reverse has been proved meaningful. A class of reversible cyclic codes is proved to have minimum distance greater than the BCH bound and it has been shown that reverse-sequence iterative decoding is the most natural technique to be coupled with the forward-sequence iterative decoding to decode this class of reversible codes beyond the BCH bound.

†This work was supported by the National Science Foundation under Grant No. GK-2339, and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199.

For cyclic codes with minimum distance greater than their BCH bound, a scheme called syndrome transformation decoding is proposed for decoding the codes beyond their BCH bound. With this scheme, the possible adaption of these codes to be used in compound channels to correct random errors, as well as burst errors, is shown in an example.

Based on the concept of syndrome transformation decoding, cyclic codes for single and multiple solid-burst-error correction have been constructed, including a simple decoding procedure.

Kenneth Tzeng

11.2.2. On Gilbert Burst-Error-Correcting Codes

The Gilbert code [1] (of density two) of length n = pq, where p and q (p < q) are relatively prime integers, is the set of binary sequences $x_0, x_1, x_2, \ldots, x_{n-1}$ which satisfy the following parity checks.

$$P_s = \sum_{i=0}^{q-1} x_{ip+s} = 0 \qquad \text{for } s = 0,1,\ldots,p-1$$

$$Q_t = \sum_{i=0}^{p-1} x_{iq+t} = 0 \qquad \text{for } t = 0,1,\ldots,q-1$$

$$P = P_0 P_1 \cdots P_{p-1}$$

$$Q = Q_0 Q_1 \cdots Q_{q-1}$$

$$b_p = \left(\frac{\pi_{p-1}}{\pi_p}\right) p$$

where π_p is smallest prime divisor of p. Similarly,

$$b_q = \left(\frac{\pi_q - 1}{\pi_q}\right) q$$

where π_q is the smallest prime divisor of q.

Neumann [2] proves in his Theorem I that a Gilbert code with parameters p and q corrects all bursts of length b' where $b' = \min(b_p, b_q)$. However, it is found that statement (3) of his theorem should be modified. We have shown that the code actually corrects all bursts of length

$b = \min\{b_p, b_q, [(p+q+2)/3]\}$. When $b' \leq [(p+q+2)/3]$, we have $b = b'$, i.e., the results of Neumann are correct in this case. When $b' > [(p+q+2)/3]$, $b = [(p+q+2)/3]$ and the capability of the codes is less than that claimed by Neumann.

L. Bahl
R. T. Chien

11.2.3. Reversible Codes

As characterized by Peterson [3] and Massey [4] a reversible code is a code such that if $(a_0, a_1, \ldots, a_{n-2}, a_{n-1})$ is a code vector, then $(a_{n-1}, a_{n-2}, \ldots, a_1, a_0)$ is also a code vector. Since a cyclic code is reversible if and only if it is generated by a self-reciprocal polynomial [3,4], we consider the binary cyclic codes generated by $g(x)$, where $g(x)$ is the minimum-degree polynomial such that has among its roots $\beta^{-(d_o-1)}, \beta^{-(d_o-2)}, \beta^{-1}, \beta^1, \beta^2, \ldots, \beta^{d_o-2}, \beta^{d_o-1}$ with $d_o = 2t+1$, $\beta \neq 0$ and $\beta \in GF(2^m)$. We found the conditions for a code generated by $g(x)$ to have minimum distance at least d_o+4. A procedure to correct up to $(d_o+3)/2$ errors was also developed. This procedure is based on Berlekamp's decoding algorithm [5] and can be implemented in computer.

A procedure was also developed to decode cyclic codes beyond the BCH bound. In this process, an equation for the unknown syndromes is found, and each solution gives us a validity error polynomial if the number of errors assumed is correct.

Carlos Hartmann

11.2.4. Threshold-Decodable Codes

BCH codes are the best known class of codes for correcting large numbers of random errors. However, the complexity of its implementation leaves something to be desired. Threshold decoding using majority logic, on the other hand, shows a promising future in this area because it is usually very simple to implement. Most of the threshold-decodable codes discovered so far are only slightly inferior to BCH codes; in fact, some BCH codes have been shown to be majority decodable.

In most of the cases of threshold decoding, a set of orthogonal parity checks is formed from the parity-check matrix for each digit.

However, Rudolph [6] has described an algorithm in which the parity checks are not necessarily orthogonal. Of the work done in this area so far, there are three approaches: (1) L-step orthogonalizable codes; the most famous of these are the Reed-Muller codes, maximal-length codes, and certain classes of BCH codes; (2) construction of the parity-check matrix satisfying certain conditions, such as orthogonality, and determining the number of information symbols [6]; (3) codes by a generator polynomial whose null space satisfies the orthogonality properties [7,8].

A new approach is being investigated. A code is defined by constructing its parity check matrix such that the parity checks on each information bit are orthogonal. Furthermore, since the parity check matrix is in systematic form, the number of information symbols of the code is known without complicated calculations. The parity check matrix is of the form

$$[P \vdots I]$$

such that I is an identity matrix, and P is either the incidence matrix of a $(b,v,k,r,\lambda=1)$ balanced, incomplete, block design (BIBD)[9] or its transpose. In either case, the orthogonal properties of the P matrix is guaranteed by the nature of the BIBD. The first case yields a $(b+v,v)$ block code capable of correcting $[r/2]$ errors, while the second case yields a $(b+v,v)$ block code capable of correcting $[k/2]$ errors. Since one of the properties of the BIBD is that $b \geq v$, the first type of code has efficiency ≤ 0.5, while the second type has efficiency ≥ 0.5.

Spencer Ng

11.2.5. A New Analysis of the Golay (23,12) Code

Rather than of applicative significance, the results presented in this section have the character of an interesting formal curiosity in the algebraic theory of error-correcting codes. In fact, they establish a connection between various codes recently discovered and the Golay (23,12) code, and provide an additional explanation for the unique nature of the latter.

First, for $b_1, b_2 \in GF(2)$ and $a_1(x)$, $a_2(x)$ and $a_3(x)$ as polynomials over GF(2) modulo $(x^{\ell}+1)$, with $\ell = 2^n-1$, we construct vectors of the format

$$[a_1(x), b_1, a_2(x), b_2, a_3(x)].$$

Then, by imposing certain independencies among $a_1(x)$, $a_2(x)$, $a_3(x)$, b_1 and b_2 we obtain linear codes $\mathcal{E}_n$, such that known classes of codes (Vasil'ev linear codes [11], nonlinear distance-5 codes [10]) are obtained as shortened versions of $\mathcal{E}_n$ or of subcodes of $\mathcal{E}_n$. Moreover, letting $n = 3$, one obtains a (23,12) whose minimum distance is shown to be 7, i.e., the code is equivalent to the Golay code. As a converse of Goethals' statement [13], it is therefore proved that the Golay (23,12) linear code is an extension of the Nordstrom-Robinson (15,8) nonlinear code [12].

Finally, it is shown that the unique nature of the Golay code can be explained with the fact that the double-error-correcting single-error-detecting primitive BCH code of length ℓ is identically 0 only for $n = 3$.

F. P. Preparata

11.2.6. Finding The Roots of a Polynomial Over Finite Field

Work continues on developing methods to find roots of a polynomial over GF(p) where p is a power of 2. General results have been obtained for cases where p is an even power of 2 and the polynomial is of degrees 2, 3 and 4.

There has been special emphasis on polynomials of degree 5. Interesting results have been obtained for cases where p is a power of 4 and 6. Present work centers on generalizing these results for all fields.

A. B. Moy

11.2.7. Probabilistic Decoding

The emphasis in this effort is to analyze a decoding algorithm for binary linear codes. The algorithm, presented in a paper by J. K. Omura entitled, "A Probabilistic Decoding Algorithm for Binary Group Codes," will always converge to the maximum-likelihood error vector. Investigation into the structure of this algorithm, which employs a gradient search over binary vectors, is promising because it is easily simulated on a computer.

An alternate approach is also presented that assumes a particular minimum-weight error vector (satisfying some constraints) to be the solution; then investigates which additional constraints the assumed vector satisfies.

In applying the decoding algorithm to the extended binary Hamming Code (Hamming (8,4)), it is found that some syndromes required a great deal of computation. One of the noticeable difficulties encountered in applying the algorithm arises in changing to a new set of basis vectors. The difficulty is that certain choices of basic matrices require much more computation for convergence to the minimum-weight error vector than others. I am attempting to predetermine which basic matrix would be the best choice, specifically which basic matrix will yield the minimum-weight error vector with the least amount of computation. The choice of the vector, containing those components of the error vector not already in correspondence with the columns of the basic matrix, can also considerably shorten the computation if the ideal choice is made. With these possible reductions in computation, the application of this algorithm to longer codes may become more practical.

Richard E. Pavek

11.3. Redundancy Techniques for Reliable Computers

11.3.1. Codes for Arithmetic Processors

Arithmetic codes are designed to correct errors in computer arithmetic operations. For additions and subtractions we have the usual AN codes to which a great deal of work has been devoted. Analysis of orbit structure and the number-theoretic considerations are the principal tools in proving the multiple-error-correcting capabilities of these codes. The most recent results, including many new codes, have been published in CSL Report 417.

Seemingly easy comparison with the well known BCH codes turns out to be very difficult for a lack of field structure in AN codes. However, partial relations beyond the cyclic property have been established and the present research is directed towards proving the minimum distance in this light. Besides the minimum distance and synthesis of codes, the

decoding problem and the implementation problem in general are studied. This area is still wide open for the obvious reason that the AN codes are hard to decode.

For the multiplication errors in the fast, parallel-scheme computers, we are working on a special class of AN codes that will correct single iterative errors.

Se June Hong

11.4. Digital Systems

11.4.1. An Estimate of the Length of Diagnostic Tests

The problem of efficient recognition of binary error-free patterns has received wide attention in several contexts, particularly in connection with the diagnosis of digital systems [14-18]. Diagnosis is based on the application of stimuli (tests) to the system and the observation of the ensuing responses. Without loss of generality, we may consider single-output combinational networks.

The text data available is, as usual, conveniently modeled as a binary s×m pattern matrix M_{sm}, whose columns--all distinct--are the fault patterns and whose rows correspond to the tests. The diagnosis problem is reduced to the selection of a minimal set of rows which preserve column distinguishability (minimal test schedule). The exact solution of the problem is usually extremely complex.

Denoting by k_{min} the minimal test schedule length for a given matrix M_{sm}, if $s \geq m-1$, Gill [4] has established lower and upper bounds on k_{min}, that is, $\lceil \log_2 m \rceil \leq k_{min} \leq m-1$. Although fundamental, these two bounds, being so far apart, can hardly be a very useful guideline.

This investigation is aimed at computing one such guideline, as an upper bound on the median of k_{min} over the set of all binary matrices M_{sm}. A counting scheme is established by simply generating, for a given test-schedule length k, a set of matrices R_k' which is properly contained within the set R_k of matrices M_{sm} having at least one test-schedule of length at most k. Thus, the lower bound $\epsilon_{sm}(k)$ is obtained on the unknown distribution $\eta_{sm}(k)$ of matrices M_{sm} which have a test-schedule of length

at most k. It is shown that $\epsilon_{sm}(k)$ is lower-bounded by $\exp(p_m(k))$, whence the median $\bar{k}$ of k is upper-bounded by $\lceil k_\infty \rceil$ where $p_{ms}(k_\infty) = \ln\frac{1}{2}$. The determination of k_∞ involves the solution of a transcendental equation. However, by modifying the Newton-Raphson method, a linear algebraic iterative approximation is obtained, whose convergence is geometric. The final result is that the average $\langle k_{min} \rangle$ of k_{min} over all matrices M_{sm} satisfies the bound

$$\langle k_{min} \rangle < \lceil 2\log_2 m \rceil$$

thereby justifying the heuristic observation by Kautz [15] that usually k_{min} is closer to its lower bound that to its upper bound.

These results have been presented at the 3rd Princeton Conference on Information Sciences and Systems. A formal paper is to appear in the IEEE Transactions on Reliability.

F. P. Preparata

11.4.2. On the Delay Required to Realize Boolean Functions

The problem of the delay required to realize switching functions has received some attention in the very recent past [19]. Specifically, given a class of logical modules to be used for network synthesis, the delay (sometimes called the computation time) of a network realizing a Boolean function is the largest number of modules in any network path connecting the input terminals to the output terminal. This largest number of modules will be called the number of levels. We denote by L(n) the largest number of levels required to realize all Boolean functions of n arguments with a given class of modules. Upper and lower bounds have been obtained by rather straightforward counting arguments [19,20].

A more interesting problem, however, is the development of constructive procedures achieving a given L(n). Following the model proposed in [19], we shall consider tree-type interconnections of arbitrary two-input one-output logic gates. We then recall that if the input variables are available in true and complemented form, only AND, OR, and EXCLUSIVE-OR gates need be considered. In Ref. [19] it was shown that, for this class of modules, L(n) = n for n = 3,4; and it was conjectured that the same

statement was true for n = 5. In this paper we show that $L(n) \leq n$ for n = 5, 6, 7 and conjecture that the same holds for n = 8. This then implies that

$$L(n) \leq n+1 \text{ for } n \leq 135$$
$$L(n) \leq n+2 \text{ for } n \leq 2^{135}+135$$

and so on.

The conclusion is based on the detailed argument used to prove the following theorem:

Every Boolean function $f(x_1,x_2,x_3,x_4)$ admits one of the two following decompositions: (1) $f = \bar{x}_j f_o \vee x_j f_1$, with f_o and f_1 realizable with at most 2 levels, or (2) $f = \psi\varphi$, with ψ and φ realizable with at most 2 and 3 levels, respectively.

F. P. Preparata
D. E. Muller

11.4.3. Generation of Near-Optimal Universal Boolean Functions

The recent technological developments in the construction of logical circuits have created several questions of a system nature, upon whose solution the efficient exploitation of the new technology will depend. A generally recognized guideline has emerged: The manufacturer should supply the designer with "modules" as large as his technology allows, and the designer should realize his circuitry by simply interconnecting modules of as few distinct types as possible.

This guideline immediately results in the desirability of a module which may be used to simulate any arbitrary Boolean function $f(x_1,\ldots,x_n)$ of n variables by appropriately connecting its inputs to wires which carry the signals x_i, their complements, or the biases 0, 1. Such module realizes a Universal Boolean Function $U(z_1,\ldots,z_m)$, i.e., U is universal for $n \geq 1$ if it realizes all Boolean functions $f(x_1,\ldots,x_n)$ by substituting for each z_j a variable of the set $(0,1,x_1,\ldots,x_n,\bar{x}_1,\ldots,\bar{x}_n)$. Related problems have been studied by other authors [21-25].

Our criterion of simplicity of $U(z_1\ldots,z_m)$ is the number m(i.e., the number of input terminals), which has been shown [21] to be lower-bounded as

$$m \geq \frac{2^n}{\log_2(2n+2)} \quad . \tag{1}$$

Let us consider the set R of the 2^n minterms in the canonical disjunctive expansion of $f(x_1,\ldots,x_n)$. We try to partition R in blocks $R_1,\ldots,R_t$. Each R_i is a block of minterms with the very special property that for each nonempty proper subset S of the minterms in R_i, there is a variable x_j such that x_j appears uncomplemented in just the terms of R_i in S, or else it appears complemented in just the terms of R_i in S. Then we construct the universal function $U = z_{n+1}R_1 \vee z_{n+2}R_2 \vee \ldots \vee z_{n+t}R_t$. In using this universal function to realize $f(x_1,\ldots,x_n)$, we always assign to the first n variables $z_1,\ldots,z_n$ the corresponding variables $x_1,\ldots x_n$ as they appear in the blocks R_i. The terms in each R_i may be selected by choice of z_{n+i} so as to be just those in the function f (which is possible by virtue of the previously stated property). Since we wish to minimize $m = t+n$, and hence t, it is desirable to make the average number of minterms in the unions as large as possible.

Representing the blocks as (sXn) binary matrices B, in which entries correspond to variables (1 for true, 0 for complemented form) and rows to minterms, the special property of the R_i may now be expressed by saying that each binary s-tuple different from 00---0 and 11---1, must appear in true or complemented form as a column of B. Clearly $s = [\log(n+1)]+1$.

We first notice that if B meets the just stated condition, then the matrix obtained by adding to each row of B a constant vector $\underset{\sim}{v}$ also meets this condition. Hence we must find a set $V \equiv \{\underset{\sim}{v}\}$ of vectors $\underset{\sim}{v}$ such that the resulting matrices contain no common row. This set is found to be a subspace of the n-dimensional vector space having dimension $(n-[\log_2 s])$.

When $s = 2^r$, a power of 2, the members of V determine a complete partition of the 2^n minterms. When $s \neq 2^r$, then a subset of the minterms are not allocated to blocks by members of V (specifically, all the members of some cosets of V). The procedure, however, can be applied recursively with a minor modification until a complete partition is obtained. The number $N(n) = t$ of blocks formed is given by the following recursion formula

$$N(n) = 2^{n-r}+(2^{r}-s)2^{x-1-r}N[n-(s-1)]. \qquad (2)$$

Application of this formula indicates that the discrepancy between N(n) and the lower bound Eq.(1) decreases for increasing n: This is due to the more general result that N(n) is asymptotically converging to the lower-bound. Therefore, we were able to show the asymptotic optimality of the generative procedure. For $n \geq 1$, although not optimal, this method provides the best known values of m.

Interesting features of this method are, besides those already mentioned: (1) the simplicity of generation of the Universal Boolean function for any $n \geq 1$; (2) the extreme simplicity in specifying the terminal interconnection in order to realize an arbitrary Boolean function $f(x_1,\ldots,x_n)$.

F. P. Preparata
D. E. Muller

11.5. Information Retrieval

11.5.1. Time-Shared Information Retrieval

Recently there has been a great deal of interest in computer-based information-retrieval systems with a large number of remote terminals. A fundamental distinction between single- and multiple-terminal systems is the problem of queuing and the consequent increase in response time. In multiple-terminal systems the response time depends not only on hardware and file organization, as in the single-terminal case, but also on the scheduling among the various terminals and the number of terminals.

Two simple models of a coordinate-indexed retrieval system have been examined. In the first, queries are processed one at a time. As an approximation for fast machines, CPU time is neglected in comparison with file-access time. The latter may be estimated from a consideration of the average characteristics of the data base and the mechanical properties of the file. It is assumed that each query has some average processing time equal to the average file-access time and that the user's reaction time is exponentially distributed, so that the first model falls within the scope of the machine-interference problem of queuing theory. For this application to information retrieval, the results show that the average response time is on the order of the product of the number of consoles with the processing

time. Furthermore, the average number of consoles waiting is a large fraction of the total number of consoles. This fact indicates that batch processing should improve the response time.

In the second model, queries are processed in batches of k at a time. No simple queuing model is available for this case. However, for certain ranges of the various parameters, an upper bound for the response time can be derived. The bound is obtained by considering a particular priority rule which reduces the queuing situation to a case that is easy to analyze. This bound depends on the batch processing time which must be estimated for the particular file under consideration. The advantage of batch processing is, of course, that with proper scheduling, k search requests can be satisfied in a shorter average time than k requests can be handled separately.

The two models have been compared, and some relations between the search time, number of consoles, and the mean and variance of the response time have been obtained. The results indicate that for systems with many terminals, batch processing, instead of one-at-a-time processing, can give considerable improvements in response time. Even so, the response time is well out of the "real-time" range. On the basis of these simple models, one must conclude that for time-shared information retrieval the real problem is the reduction of file-search time. Some work is in progress on this problem.

T. Cordaro

11.5.2. Document Retrieval Based on Semantic and Bibliographic Indexing

In the previous progress report, we outlined a document-storage and retrieval scheme based both on semantic and citation indexing, in order to take advantage of the consensus of these two independent structures.

Preliminary manual experiments have been performed on a small data base (approximately 120 documents) in the field of "Computer-Aided Instruction", in order to verify the reasonableness of the working hypothesis. The results have been sufficiently positive to encourage setting up a more extensive data base for more complex experimentation, especially with reference to retrieval strategies.

The storage format is based on an 8-digit coding of documents (which is expected to be unambiguous) and each document is stored as follows: document code, basic bibliographic data (author(s) and title), document codes of the cited references, and descriptors.

The document collection has been chosen in the field of Computers, and the journals currently being examined are the Journal of ACM, IEEE Transactions on Computers, and the IEEE Transactions on Information Theory. The distribution of the cited references will indicate in which direction the collection of journals should be augmented in order to enrich the citation network.

Elizabeth R. Lyman
Franco P. Preparata

11.6. Stochastic Systems

11.6.1. Dynamical Representation of Markov Processes of the Separable Class

A Markov process $x(t)$ is of the separable class if it satisfies the property

$$E\{x(t+\tau)|x(t)\} = \rho(\tau)[x-m_o]+m_o$$

where $\rho(\tau)$ is the normalized autocorrelation function of the process, and m_o is its mean. The purpose of this work was to show that a stationary Markov process of the separable class with $\rho(\tau) = e^{-a|\tau|}$ may be represented by the stochastic differential equation

$$dx(t) = -a(x-m_o)dt + g(x)dw(t)$$

where $w(t)$ is a Wiener process, and $g(x)$ is a nonlinearity which determines, or is determined by, the first-order probability density of the process. The properties of such processes were investigated, and it was shown that a large class of stationary Markov processes may be transformed into separable Markov processes by simple zero-memory nonlinearities. The importance of the representation is in its application to optimum nonlinear filtering and prediction.

The results were also extended to the multidimensional case, which includes the generalized separable process $y(t)$ satisfying the property

$$E\{y(t+\tau)\,|\,y(t)=y\} = \sum_{i=1}^{n} \rho_i(\tau) f_i(y)$$

Several examples were considered as illustrations.

A. H. Haddad

11.6.2. Approximate Design of Control Systems in the Presence of Noise[†]

Two recent results in the approximate design of deterministic control systems were generalized to include the cases of noisy measurements and noisy inputs.

The first [26] is concerned with the design of control systems with uncertain parameters and noisy measurements which was considered in Ref. [27] for the noiseless case. An approximate solution is proposed which separates the estimators from the controller. The estimates of the uncertain parameters are obtained from a first-order approximation of the deviation of the trajectory from the nominal trajectory. An example was considered and simulated, and it illustrates the principal property of the approximate design: For low noise, the system approaches the noiseless design of Ref.[27], and, for the high-noise case, the system approaches the open-loop design.

The second extends the decoupling method discussed in Ref.[28] to systems with white-noise inputs and noisy measurements. The decoupling results obtained for the filters are similar to those obtained for the optimal control.

A. H. Haddad
P. V. Kokotović
J. B. Cruz, Jr.

11.7. Channel and Source Coding

11.7.1. Time-Continuous Channels

Channels of the following description are being studied: inputs

[†]This work was supported by the Air Force Office of Scientific Research under Grant No. AFOSR-68-1579, and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199.

are square-integrable functions of norm $\leq ST$ (where T is the period of operation of the channel); the inputs are transformed by an integral operator (a linear filter) and Gaussian noise of known spectral density is added. The object of the investigation is to find representations, principally in terms of reproducing-kernel Hilbert spaces, for stochastic inputs to such a channel, and to determine the capacity of the channel for such inputs, thereby clarifying the characteristics of optimal coding for such channels.

John K. Moore

11.7.2. Text Compression

We have been working on a method of compacting written text in digital computers using statistical methods. We started out by examining English text and word frequencies in small subtexts of a large text. After doing this, we decided that it was reasonable to break up texts into small sections and encode them separately, thus, taking advantage of the local structure which we observed in small text sections.

In order to take complete advantage of this local structure, the encoding of each small text section has to be different. This means there has to be an encoding of a description of the encoding of each text section. The encoding of each of these takes up a considerable amount of space, more space than was saved by coding in small sections.

The solution of this problem depends on the method of encoding of the small-text sections. The way we encoded them was to consider them as having been produced by a zero-memory source which was most likely to have produced this text segment. We then used the Huffman encoding of the messages of this source to encode the text. The description of the encoding of a small text is thus the description of a zero-memory source.

A description of a zero-memory source is rather large; however, if all the zero-memory sources of many small-text sections are described given the knowledge of a description of the zero-memory source that would have produced the sum of all the sections, then the size of each description is reduced considerably.

We can calculate the size of the description of the zero-memory source description of the small text section. Let T be a large section of text of length N. Let the number of different words in T, $\{m_i\}$, be D. Let the number of occurrences of m_i in T be F_i. Let T' be an arbitrarily chosen subset of T of length M. Let f_i be the number of occurrences of m_i in T'. We need to know $P(f_i=g)$, i.e., the probability that for a fixed i, $f_i = g$. A straightforward combinatorial argument yields

$$P_{ig} \triangleq P(f_i = g) = \binom{F_i}{g} A_{ig}$$

where

$$A_{ig} = \left(\frac{N-M}{N}\right)\left(\frac{N-M-1}{N-1}\right)\cdots\left(\frac{N-M-F_i+g+1}{N-F_i+g+1}\right)\left(\frac{M}{N-F_i+g}\right)\left(\frac{M-1}{N-F_i+g-1}\right)\cdots\left(\frac{N-g+1}{N-F_i+1}\right)$$

Therefore, the number of bits required to express the value of f_i is

$$\sum_{g=0}^{f_i} P_{ig} \log_2(1/P_{ig})$$

Thus, to tell what all the various f_i are takes

$$\sum_{i=1}^{D} \left[\sum_{g=0}^{F_i} P_{ig} \log_2(1/P_{ig}) \right]$$

bits. That is, this is the expected value of the description of a zero-memory source corresponding to a text of length M given that it is a subtext of a text of length N.

The words of the text section are chosen to be the messages of the zero-memory source because of the following reasons:

1. Words are convenient to work with for many reasons; one of these being that it is easy to describe to a computer what a word is.

2. Using letters or syllables for messages would increase the size of the Huffman encoding of the text greatly.

3. Using multiword messages would mean a great increase in the size of the zero-memory source description.

These reasons are not sufficient to show that using words as messages is optimal and it is, in fact, almost certain to be not optimal. However, it is probable that for most coding applications that some or all of these reasons justify the use of words as messages.

An essential tool for this investigation is an efficient program designed to obtain the compact encoding of a zero-memory source (Huffman encoding). This program, named HUFF, whose interest goes beyond the objectives of the present project, is briefly described below.

Huffman Encoding Program (HUFF). The way HUFF operates is that it is given the names of two integer arrays L1 and L2 and the integer N. To encode a set of n messages with the frequency of the ith message being f_i, set N = n and set L1(i) = f_i for $1 \leq i \leq n$ and call HUFF. When HUFF returns the binary encoding of the ith message is contained in the low order K_i bits of L1(i) where K_i = L2(i). (The L2 array is automatically set by HUFF during the encoding process.)

The encoding done by HUFF is a three step process.

STEP 1 - SORTING: The sorting done is a two step process.

1. The L1 list of frequencies is scanned and the different members are each placed in one of 101 different lists. The list that a member is placed in is determined by the following criteria:
   a. If L1(i) < 101 then the ith member (i.e., L1(i)) is placed in the list which has as its number L1(i).
   b. If L1(i) $\geq$ 101 then the ith member is placed in list 101. The time taken by this step of the sorting is linear with respect to the length of the L1 array (i.e., N.)
2. The 101th list of frequencies is now sorted according to frequency using a conventional "bubble" sort. Notice that if the 101 list are placed end to end, the large list produced is a list of all messages sorted according to frequency. (This is not actually done but the set of lists can be thought of in this way.) Also, notice that if the number of frequencies greater than 100 is small with respect to the total number of frequencies then the sorting time is essentially linear with respect to N.

STEP 2 - TREE FORMATION: The tree formation is done as follows. The set of 101 lists is considered as one large list and the two smallest frequency elements are found and combined into one "node". This node is then placed in the large list according to its frequency. The components of this node are recorded along with its frequency. (The frequency of a node is the sum of the frequencies of its components.) This process continues until there

is only one node left. Since the search process for a node is a constant with respect to the number of nodes, if the number of frequency elements with frequency greater than 100 is small, the tree-formation process is linear with respect to N because the number of nodes is N-1.

STEP 3 - CODE FORMATION: The code formation is done by searching along each branch of the binary tree formed by the last step and assigning code words at each terminal node in the standard way that is used in Huffman encoding. In HUFF this is done by simulating within HUFF a "recursive subroutine", RS. This RS is given a code length corresponding to the depth into the binary tree, the code word so far built up and the pointer to the present node being considered.

The time it takes to produce a bit of code is constant, so the time that it takes to do the last step is linear with respect to M, the total number of bits in the code words.

The outstanding problem we are presently concerned with is to devise a reasonably simple strategy for partitioning the text into small sections in order to obtain a near-optimal encoding.

Robert Cullum

11.7.3. Channel Modeling

In information storage, transmission, and retrieval, three primary error processes corrupt data: (1) Burst errors
(2) Random errors
(3) Periodic errors.

Much effort has been expended to determine appropriate mathematical models both to assist in evaluating proposed error-protection coding schemes and to extend present knowledge of the various error processes and their interaction. One of the most tractable models is one proposed by Elliot (using data gathered from the switched telephone network), in which P(m,n) curves (probability of m errors in n transmitted digits) are modeled by a sum of component renewal channels. For each value of n (code length) considered, there is a characteristic "hump and tail" in the P(m,n) curve, the peak of the hump occurring at approximately n/2. The fact that the location of the peak and the wideness of the hump appear to be linearly related to n, suggests that the acting error process

has a periodic component. Consequently, P(m,n) curves for error sequences produced by passing periodic waveforms of zeros and ones through a binary symmetric channel (BSC) are being investigated to determine if such a pseudo-periodic channel can be satisfactorily used as one of the component channels to model the switched telephone network. Present indications are that a periodic sequence of six ones followed by six zeros (12 bit period) passed through a BSC with an error probability of 0.2, produces an error sequence the P(m,n) curve for which has a "hump" with the correct wideness, peaking at approximately n/2. Present efforts are directed at developing computer programs to calculate P(m,n) for varying duty cycle of the periodic sequence and error probability for the BSC. If these further investigations prove fruitful, it may be possible to attribute the hump and tail of the P(m,n) curves for the switched telephone network to errors caused by power or signaling frequencies.

Paul Davis

11.8. Communication Technology

11.8.1. Resolver-Type FM Modulators

Serrasoid generators using polarity-switched integrators of the type discussed in the previous progress report [30] have been the subject of an experimental program of design and construction. Emphasis was given to design parameters that would find application in the frequency-modulation and to a determination of the design limitations imposed by the state of the present development of semiconductor devices and by reasonable cost considerations.

A report is now in preparation that concludes that this approach is not competitive in cost and simplicity with present-day methods of achieving frequency modulation to the same degree of accuracy.

D. H. Cooper
R. J. Kubala

11.9. References

1. E. N. Gilbert, "A Problem in Binary Encoding", Proc. 1960 Symp. in Appl. Math. 10, pp. 291-297.

2. P. G. Neumann, "A Note on Gilbert Burst-Correcting Codes," IEEE Trans. on Info. Theory IT-11, pp. 377-384, July, 1965.

3. W. W. Peterson, Error Correcting Codes, M.I.T. Press, John Wiley, N. Y., 1961.

4. J. L. Massey, "Reversible Codes," Information and Control, 7, pp. 369-380, 1964.

5. E. R. Berlekamp, Algebraic Coding Theory, McGraw-Hill, New York, 1968.

6. L. D. Rudolph, "A Class of Majority Decodable Codes", IEEE Trans. on Info. Theory, pp. 305-307, April, 1967.

7. E. J. Weldon, Jr., "Difference-Set Cyclic Codes", B.S.T.J., 45, pp. 1045-1055, Sept., 1966.

8. D. K. Chow, "A Geometric Approach to Coding Theory with Application to Information Retrieval", CSL Report R-368, Oct., 1967.

9. H. B. Mann, "Analysis and Design of Experiments", Dover Publications, 1969.

10. F. P. Preparata, "A Class of Optimum Nonlinear Double-Error-Correcting Codes", Information and Control 13, pp. 378-400, 1968.

11. Yu. L. Vasil'ev, "O negruppovikh plotno upakovannykh kodakh (On Nongroup Close Packed Codes)", Problemy Kibernetiki 8, pp. 337-339, 1962.

12. A. W. Nordstrom and J. P. Robinson, "An Optimum Nonlinear Code", Information and Control 11, pp. 613-616, 1967.

13. J. M. Goethals, "On the Golay Perfect Binary Code," Report R93, M.B.L.E. Laboratoire de Recherches, Brussels, Belgium, Oct. 1968.

14. R. W. Downing, J. S. Nowak, and L. S. Tuomenoksa, "No. 1 ESS Maintenance Plan," Bell Syst. Tech. Jour. 43, pp. 1961-2020, Sept. 1964.

15. W. H. Kautz, "Fault Testing and Diagnosis in Combinational Digital Circuits", IEEE Trans. C-17, pp. 352-366, April, 1968.

16. H. Y. Chang, "An Algorithm for Selecting an Optimum Set of Diagnostic Tests", IEEE Trans. on Electronic Computers, EC-14, pp. 706-711, October, 1965.

17. Arthur Gill, "Minimum-Scan Pattern Recognition", IRE Trans. Information Theory IT-5, pp. 52-58, June, 1959.

18. E. C. Kiekeman, A. Glovazky, and E. J. McCluskey, "Determination of Redundancies in a Set of Patterns", IRE Trans. Information Theory (Correspondence) IT-3, p. 167, June, 1957.

19. B. Elspas, W. H. Kautz, H. S. Stone, "Properties of Modular Multi-functional Computer Networks," Final Report, Project 4641 (AFCRL) Stanford Research Institute, Menlo Park, California (November, 1968).

20. P. S. Spira, "On the Time Necessary to Compute Switching Functions," IBM Research Report RC 2247 (#11116), Yorktown Heights, N.Y., October 14, 1968.

21. B. Elspas, W. H. Kautz, H. S. Stone, "Properties of Cellular Arrays for Logic and Storage," Stanford Research Institute Report AFCRL-68-0005, Nov. 1967, pp. 27-36.

22. B. Elspas, J. Goldberg, C. L. Jackson, W. H. Kautz, H. S. Stone, "Properties of Cellular Arrays for Logic and Storage", Stanford Research Institute Report, AFCRL-67-0463, July 1967, pp. 59-81.

23. Y. N. Patt, "A Complex Logic Module for the Synthesis of Combinational Switching Circuits," A.F.I.P.S. Conf. Proc. 30, 1967, pp. 699-706.

24. W. Frank King III, "The Synthesis of Multipurpose Logic Devices," IEEE Conf. Rec. 1966, Seventh Annual Symposium on Switching and Automata Theory, pp. 227-235.

25. D. C. Forslund and R. Waxman, "The Universal Logic Block (U.L.B.) and its Application to Logic Design," IEEE Conf. Rec. 1966, Seventh Annual Symposium on Switching and Automata Theory, pp. 236-250.

26. A. H. Haddad, P. V. Kokotović, and J. B. Cruz, Jr., "Design of Control Systems with Uncertain Parameters", The Third Ann. Princeton Conference on Information Sciences and Systems, March 27-28, 1969.

27. P. V. Kokotović, J. B. Cruz, Jr., J. E. Heller, and P. Sannuti, "Synthesis of Optimally Sensitive Systems", Proc. IEEE 56, 1318-1324, (August 1968).

28. P. V. Kokotović, W. R. Perkins, J. B. Cruz, Jr., and G. D'Ans, "ε-Coupling Method for Near Optimum Design of Large Scale Linear Systems", to appear in Proc. IEE (London), July, 1969.

29. E. O. Elliott, A Model for the Switched Telephone Network for Data Communications, Bell Syst. Tech. J. pp. 89-109, January 1965.

30. Progress Report for March through August, 1968, Coordinated Science Laboratory, Univ. of Ill., (September 30, 1968), pp. 281-289.

T. N. Trick M. E. Mokari-Bolhassan J. Vlach
W. Mayeda S. Toida N. Wax

12.1. Computer-Aided Design of Distributed Networks

A computer-aided-design approach is being developed for the broad-band matching of complex generator and load impedances to distributed-active networks. If a reliable and economical method can be developed, it would be useful, for example, in the broad-band matching of an antenna to an active, lumped-distributed system in order to achieve optimum low-noise performance in the case of a receiver, or maximum power transfer in the case of a transmitter. In addition, such a program could be used in the design of distributed filters.

A network analysis program has been written based on the theory of the transfer-scattering parameters. This program will analyze any two-port cascade of lumped-distributed elements, including active elements such as transistors, and any complex terminations are allowed. The program also checks the stability of the two port. Presently a modified version of Rosenbrock's optimization procedure is being used to minimize the function

$$F = \sum_{i=1}^{n} \{a(|s_{11}^i|^2 + |s_{22}^i|^2) + b(|s_{21}^i|^2 - |s_{21}^{avg}|^2)^d\},$$

where s_{11} and s_{22} are the reflection losses and $|s_{21}^i|^2 - |s_{21}^{avg}|^2$ is the deviation of the power gain from some predetermined average value s_{21}^{avg}. The quantities a, b, and d are weighting factors. The summation is taken with respect to frequency, and the index i indicates one of the finite number of frequencies ω_i, i=1,2,...,n, where $\omega_i \in \Omega_o$, and Ω_o is the frequency range of interest.

To date our results have been very encouraging. Figure 12.1 indicates a typical matching problem. The network consists of a cascade connection of lossless series lines, stubs, and a transistor. It is desired to find the length and characteristic impedance of each line such

†This work was supported by the Air Force Office of Scientific Research under Grant No. AFOSR 931-67 and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB-07-67-C-0199.

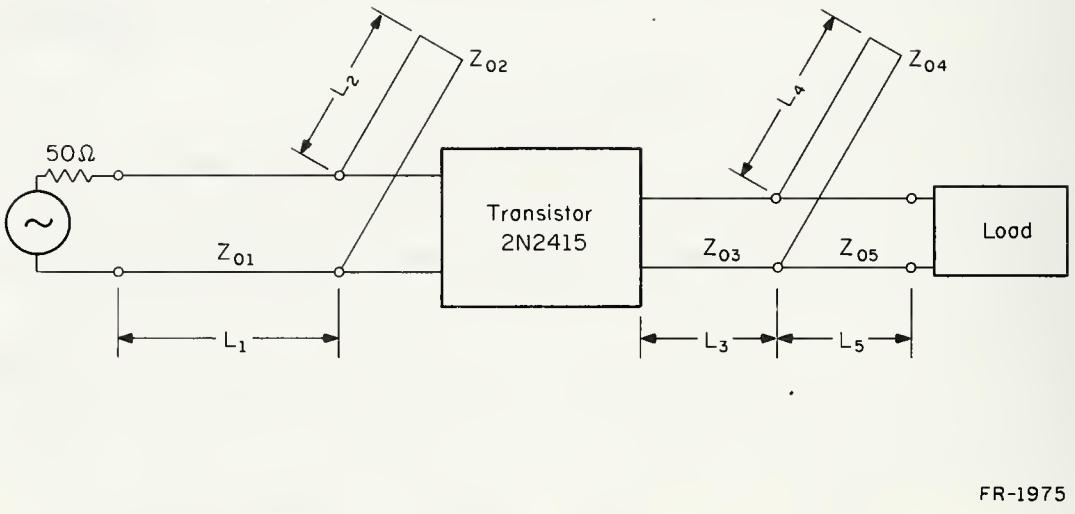


Fig. 12.1. Typical matching network.

that the power delivered to the load is optimum in some sense over the frequency range 150MHz-300MHz. In the first case we assumed a 50Ω resistive load. Figure 12.2 indicates some of our results as a function of the weighting factors. Note that our initial parameter values gave a very poor result. Figure 12.3 compares one of the computer-aided designs to an experimental result. This comparison is very favorable since they only differ by approximately 0.5 dB. It was felt that the primary source of error was in the measurement of the transistor parameters. Figure 12.4, which gives a comparison of the experimental and computed designs without the transistor, supports this hypothesis. The experimental loss is slightly greater than the computed loss, since the computer program was written for lossless lines.

In the second example, the load impedance was a slot antenna, whose impedance over the above frequency range is illustrated in Fig. 12.5. The tremendous variation in the real and imaginary parts of this load over the frequency range of interest seems to imply that a good match would be impossible; yet Figs. 12.6 and 12.7 indicate that this is not the case. The scale is exaggerated, but note that the gain variation is only about 3.5 dB over the entire frequency range.

This work was performed in conjunction with the University of Illinois Antenna Laboratory which conducted the experimental measurements. Presently other optimization techniques are being investigated.

T. Trick
J. Vlach
M. Mokari-Bolhassan

12.2. Linear Graphs and System Diagnosis

A simple proof of Tutte's theorem on realizability of cut-set (circuit) matrices has been found. This theorem gives one of the very important properties of linear graphs, which has many applications, especially to traffic flow problems and data (and voice) communication systems.

Presently there are no basic mathematical theorems in the area of system diagnosis, such as computer diagnosis. Since a linear graph is a

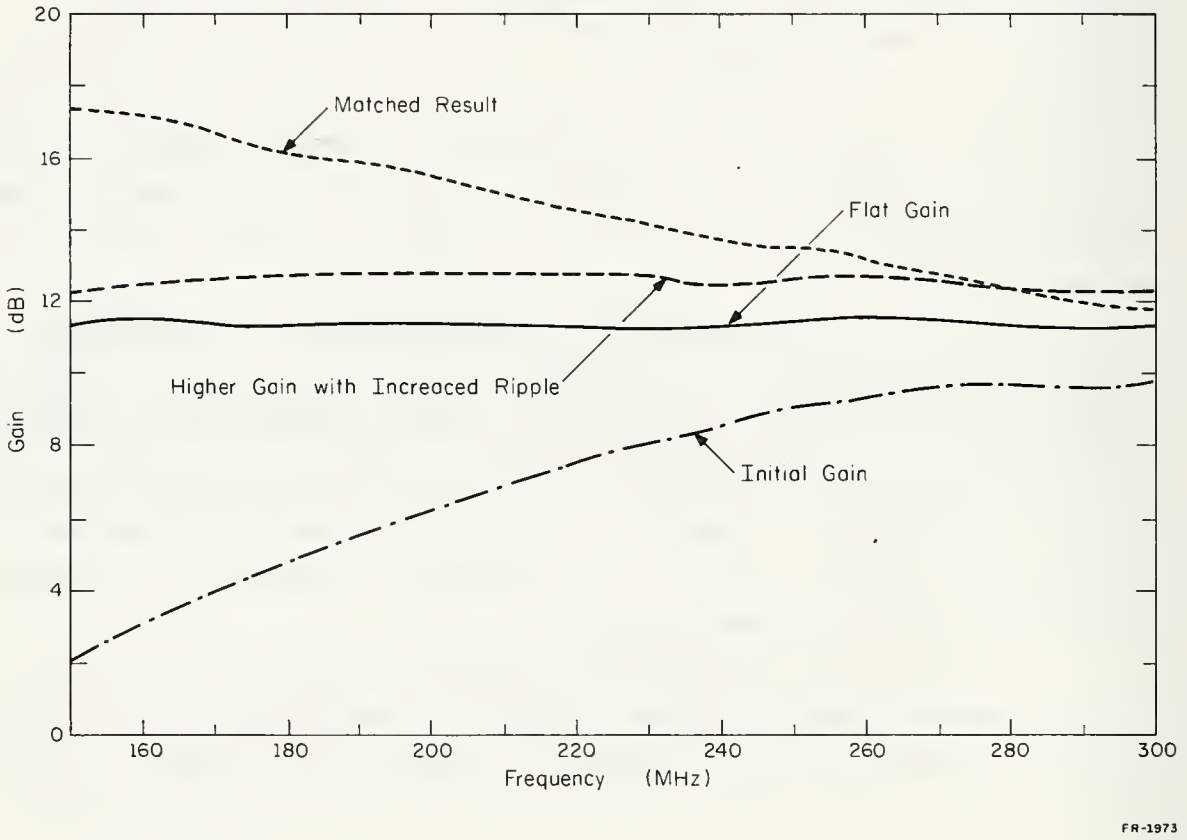


Fig. 12.2. Gain as a function of the weighting factors.

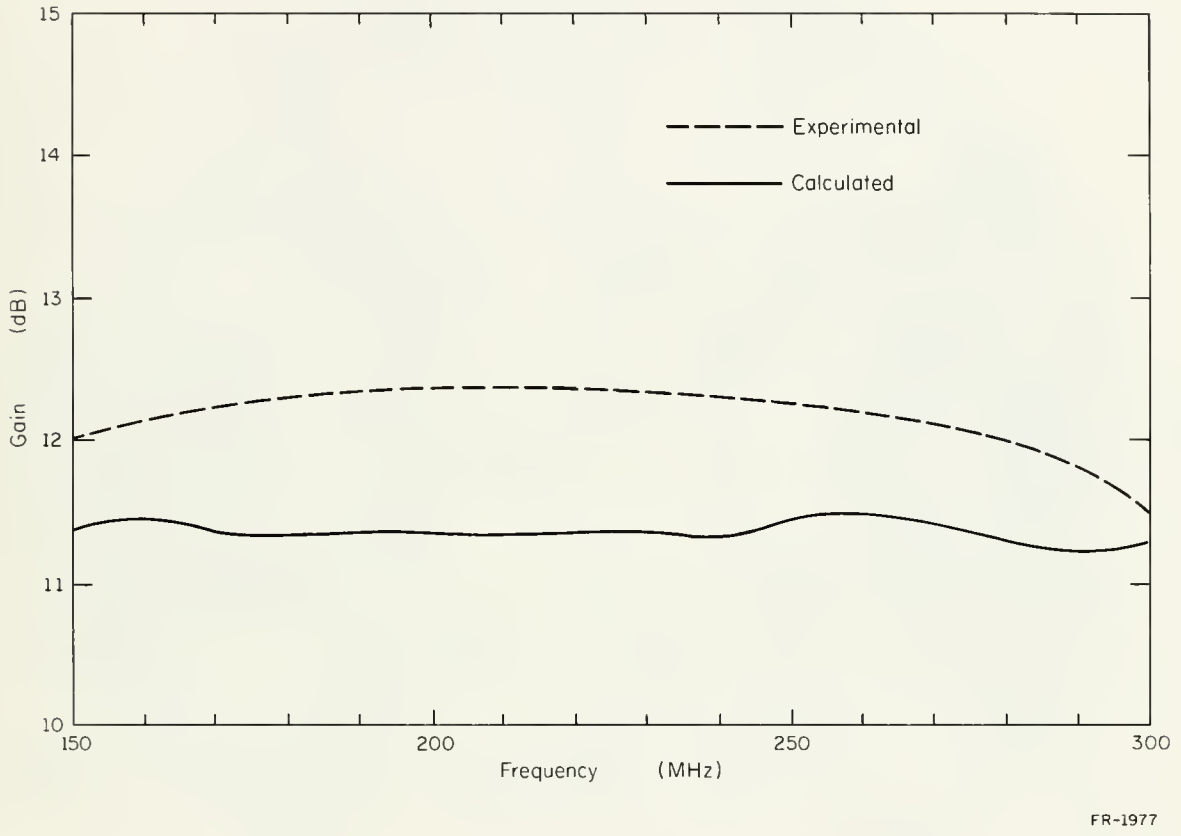


Fig. 12.3. Experimental result.

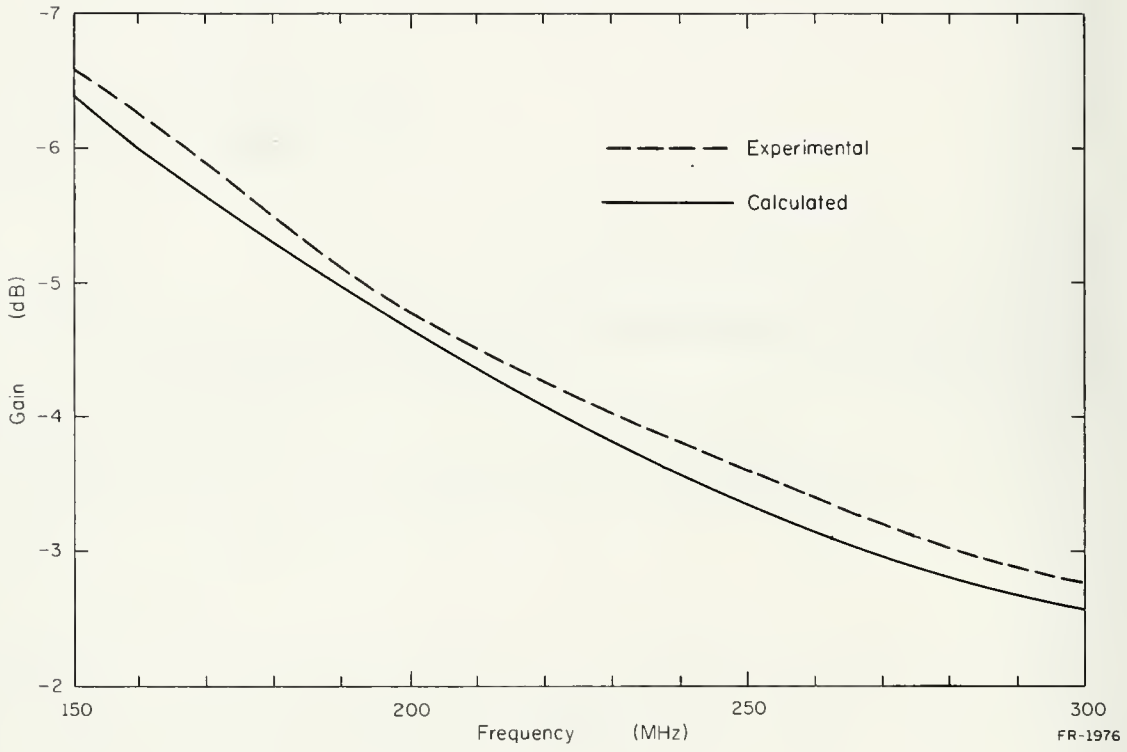


Fig. 12.4. Experimental result without the transistor.

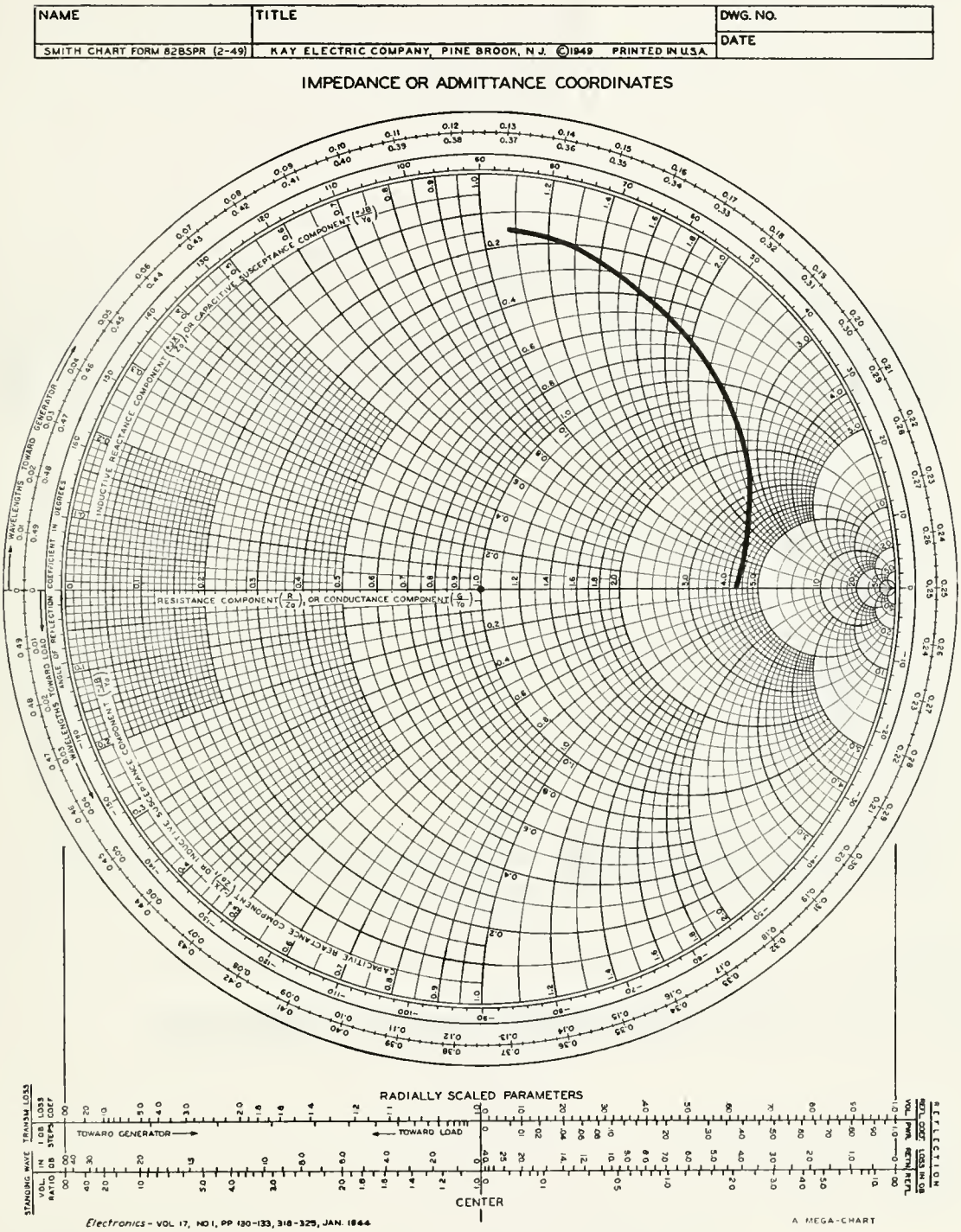


Fig. 12.5. Impedance of a slot antenna vs frequency.

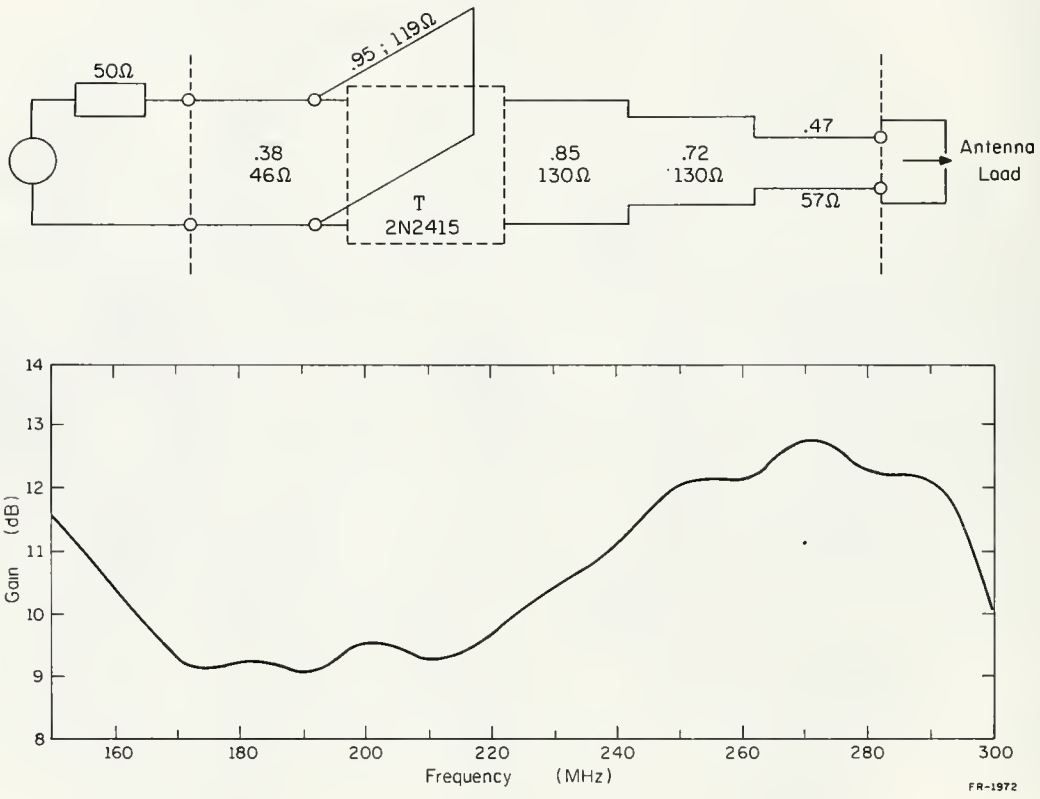


Fig. 12.6. Gain with matching network and slot antenna load.

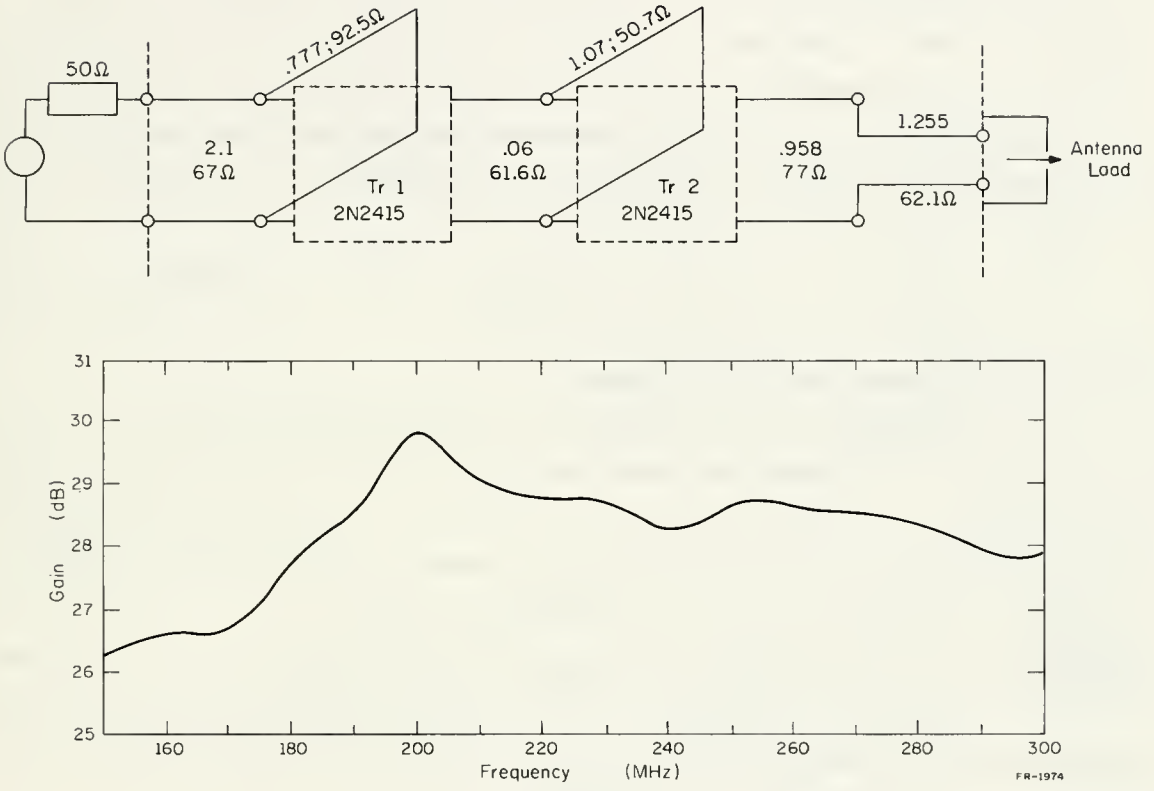


Fig. 12.7. Two-stage gain with slot antenna load.

reasonable tool to represent systems whose performance depend not only on kinds of components, but also on relative locations of components, we are studying system diagnosis from the topological point of view. One known result in the area of system diagnosis is a necessary condition for the 1 distinguishability of a system when test points are employed. Another is a necessary condition for the 1-(ij) distinguishability when only test gates are used. A problem we are studying now is to find conditions for 1-(ij) distinguishability when test signals are employed. Also the study of conditions for the 1 distinguishability of systems when combinations of test points, test gates, and test signals are used, is in progress.

W. Mayeda

12.3. Biplanar and Euler Graphs

A graph is called biplanar if it can be partitioned into two planar graphs. Although necessary and sufficient conditions for a graph to be biplanar are very difficult to obtain, it is not so difficult to obtain some relations between biplanar graphs and the degree of their vertices. Our study led to the following results:

Theorem 1. If every vertex in a graph has degree less than 5, then the graph is biplanar.

Theorem 2. If every vertex in a graph has degree greater than 11, then the graph is not biplanar.

Theorem 3. There is a biplanar graph and a nonbiplanar graph among every regular graph of degree 7.

In addition a study of the properties of Euler graphs was made. A graph is called Euler if every vertex has even degree. The following theorems were proved.

Theorem 1. The number of distinct circuits containing an edge in an Euler graph is odd.

Theorem 2. The number of paths between any two vertices in an Euler graph is even.

Theorem 3. The number of edges in a cut set in an Euler graph is even.

More details of these results and others on the characterization of biplanar graphs can be found in CSL Report R-415.

S. Toida

12.4. Nonlinear Oscillations

A study of the free-running multivibrator as a nonlinear generator of almost-discontinuous oscillations has been completed and is being prepared for publication. One manuscript, giving a general mathematical background, has been submitted for publication. Another is in the final-draft stage, and the third manuscript is in preparation.

N. Wax

J. B. Cruz, Jr.
C. I. Chen
G. D'Ans
J. Davis
J. E. Heller
R. Hecht
M. Jamshidi
P. Kokotović
J. Medanić
W. R. Perkins
M. Rice
G. Singh

13.1. Introduction

This group is concerned with the analytical, computational, and simulation aspects of control-system analysis and design. During this reporting period, special attention has been given to the development of computationally attractive approaches to design of near-optimum large-scale systems. Furthermore, several additional results concerning minimax algorithms and feedback strategies for differential games have been developed. Suboptimal filtering, parameter sensitivity of distributed systems, and adaptive systems have also received attention. Two applications of optimal control theory are being studied in detail, the first involving a traffic-control problem, the second involving a steel-rolling mill.

J. B. Cruz, Jr.
W. R. Perkins
P. Kokotović

13.2. Feedback Design of Large-Scale Linear Systems

In order to make analysis and design of large-scale systems practical, system analysts and designers try to split a large system problem into a set of simpler subsystem problems. Examples of this approach in process control, space guidance and other fields of control engineering are numerous. Consider how often different subsystems for regulation of temperature, pressure, and flow are designed separately in spite of their connection through a chemical plant, or how frequently each axis of a three-axis spacecraft attitude control system is designed separately. A reason for this decoupling practice is clear from a simple example. For a coupled 12th-order linear system, 78 scalar, coupled, nonlinear equations must be

†This work was supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR 68-1579 and by the National Science Foundation under Grant NSF GK-3893 and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB-07-67-C-0199.

solved. If it is possible to design this system as three uncoupled 4th-order systems, then three uncoupled sets of only 10 scalar equations are to be solved. The reduction of a computation effort is even more impressive if higher-order systems are decoupled in several subsystems.

The reduction of computation effort has become one of the important issues of a modern system design, since very often the amount of computation is the deciding factor on whether a large system can be designed and implemented, or not. Limited computation speed, limited word length, or limited memory are among the most critical computer constraints, especially in real-time control and estimation. In these applications, the designer is forced to neglect couplings in a coupled system and thus to reduce the problem to fit the computer constraints. Inevitably, there is a sacrifice of the performance achieved. Unfortunately, in the conventional approach that completely neglects coupling, this sacrifice is often too large, and the performance of the designed system may not be satisfactory.

An important idea in the development is the notion of $\mathcal{E}$-coupling. A large system is said to be $\mathcal{E}$-coupled if it splits into several independent subsystems when a scalar parameter $\mathcal{E}$ is zero. The partitioned form of a linear $\mathcal{E}$-coupled system consisting of two subsystems is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_1 & \mathcal{E}A_{12} \\ \mathcal{E}A_{21} & A_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 & \mathcal{E}B_{12} \\ \mathcal{E}B_{21} & B_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} ,$$

where x_1 and x_2 are n_1- and n_2-dimensional substates, with $n_1 + n_2 = n$, and u_1 and u_2 are r_1- and r_2-dimensional subcontrols, with $r_1 + r_2 = r$. The submatrices A_1, A_2, A_{12}, A_{21}, B_1, B_2, B_{12}, and B_{21} do not depend on $\mathcal{E}$.

In this work, we sought a compromise between the numerical practicality of decoupled design and the performance J_∞ achieved with regulators optimal for large coupled systems. We introduced the notion of $\mathcal{E}$-coupling and developed a method for computing near-optimum regulators. The feedback matrix $M = M(t,\mathcal{E})$ of a near optimum regulator is defined as a truncated power series in the coupling parameter $\mathcal{E}$. It is shown here that this series can be obtained from decoupled subsystem calculations. In this way, the amount of computation, compared to that for obtaining the feedback matrix P of the optimum regulator, is greatly reduced, while, in view of a theorem

reported last year, an m-th order, near-optimum regulator yields a performance which approximates the optimal performance to the (2m+1)th order.

Details of this result may be found in "ε-Coupling Method for Near-Optimum Design of Large-Scale Linear Systems," Proc. IEE, May 1969.

P. Kokotović
G. D'Ans
J. B. Cruz, Jr.
W. R. Perkins

13.3. Parameter-Imbedding Design of Linear Optimal Regulators

The solution of the regulator problem for a completely controllable, linear, time-invariant plant

$$\dot{x} = Ax + Bu, \tag{1}$$

and a quadratic performance index

$$J = \tfrac{1}{2}\int_0^\infty (x'Qx + u'Ru)dt, \tag{2}$$

is known to be $u = -R^{-1}B'Kx$, where K is the positive definite root of the matrix quadratic equation

$$KA + A'K - KSK + Q = 0, \tag{3}$$

and $S = BR^{-1}B'$. Applications of this useful result encounter some difficulties. One of them is that some trials are needed in order to find an adequate choice of Q and R. The solution of the algebraic equation (3) is usually obtained as the steady state of the solution K(t) of a Riccati differential equation. It is hard to predict how long it will take an integration routine to reach the steady state for a choice of Q and R. For high-order systems the convergence $dK/dt \to 0$ may be too slow and the trials with different Q and R may require a prohibitive amount of computation. This difficulty is even greater if the matrices A and B depend on a parameter ε, and the solution K is to be obtained for a number of values of ε.

In the parameter-imbedding method proposed here, only one imbedding equation (4) is solved instead of repeated solutions of the Riccati equation.

For all $\varepsilon \in [\varepsilon_0, \varepsilon_1]$, let A, B, Q, and R be continuously differentiable functions of a parameter ε, the system (1) be completely controllable,

and R and Q be positive definite. Then the positive-definite solution K of (3) is a continuously differentiable function of $\mathcal{E}$ for $\mathcal{E}\in[\mathcal{E}_o,\mathcal{E}_1]$. Let this function be known for $\mathcal{E} = \mathcal{E}_o$. Using $K^o = K(\mathcal{E}_o)$ as the initial condition, the solution $K = K(\mathcal{E})$ of (3) can be obtained for all $\mathcal{E}\in[\mathcal{E}_o,\mathcal{E}_1]$ from the following differential equation:

$$dK/d\mathcal{E}(A-SK) + (A-SK)'dK/d\mathcal{E} = K\alpha-\alpha'K + K\zeta K - \sigma, \tag{4}$$

where $\alpha = dA/d\mathcal{E}$, $\zeta = dS/d\mathcal{E}$, $\sigma = dQ/d\mathcal{E}$.

Depending on the nature of the parameter $\mathcal{E}$, two different applications of the imbedding equation (4) are discussed below.

13.3.1. Weighting of Q and R

An approach to the selection of Q and R is to define $Q(\mathcal{E}) = \mathcal{E}\hat{Q}$, so that the value of $\mathcal{E}$ determines the relative weights of the two terms in (2). Then (4) reduces to

$$dK/d\mathcal{E}\ (A-SK) + (A-SK)'dK/d\mathcal{E} = -\hat{Q}, \tag{5}$$

since A, B, S, and $\hat{Q}$ do not depend on $\mathcal{E}$. A "satisfactory" K can be selected from the family $K = K(\mathcal{E})$.

13.3.2. Slowly-Varying Systems

It often happens that A and B are time-varying matrices but their variations are slow compared with the transient of the system (1). In this situation $\mathcal{E}$ may be introduced to represent one or more physical parameters and (4) can be used to obtain $K(\mathcal{E})$, as in the following example.

In a tension-regulator problem, the state variables are the angular velocity x_1 and the tension x_2.

The coil diameter ρ, which slowly varies from $\rho = 1$ to $\rho = 3$, strongly influences the system dynamics. For example, the coil moment of inertia increases 81 times during the process. Therefore, the optimum feedback matrix K must depend on ρ. Numerical values of a particular strip winding plant are used to clarify the computational procedure,

$$A(\mathcal{E}) = \begin{bmatrix} -.04\mathcal{E} & -.4\mathcal{E} \\ 261 & -.36 \end{bmatrix}, \quad B(\mathcal{E}) = \begin{bmatrix} .09\mathcal{E} \\ 0 \end{bmatrix},$$

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & 500 \end{bmatrix}, \quad R = 1,$$

where $\mathcal{E} = 1/\rho^2$ and therefore $\mathcal{E}\in[1/9, 1]$. Equation (3) becomes

$$b_1^2 k_{11}^2 - 2a_{11}k_{11} - 2a_{21}k_{12} = 0,$$
$$b_1^2 k_{11}k_{12} - 2_{12}k_{11} - a_{22}k_{12} - 2_{11}k_{12} - a_{21}k_{22} = 0, \qquad (6)$$
$$b_1^2 k_{12}^2 - 2a_{12}k_{12} - 2a_{22}k_{22} - q_{22} = 0,$$

where $a_{11}, a_{12}, \ldots, b_1, q_{22}, k_{11}, k_{12}, k_{22}$ are the elements of A, B, Q, and K. Upon differentiating (6) with respect to ε and substituting the numerical values, there is obtained

$$\begin{bmatrix} \frac{dk_{11}}{d\varepsilon} \\ \frac{dk_{12}}{d\varepsilon} \\ \frac{dk_{22}}{d\varepsilon} \end{bmatrix} = \begin{bmatrix} .2\varepsilon^2 k_{11} - \varepsilon & -5400 & 0 \\ .2\varepsilon^2 k_{12} + \varepsilon & .2\varepsilon^2 k_{11} + \varepsilon + 9 & -5400 \\ 0 & .2\varepsilon^2 k_{12} + 10\varepsilon & 9 \end{bmatrix}^{-1} \begin{bmatrix} .2\varepsilon k_{11}^2 + k_{11} \\ .2\varepsilon k_{11}k_{12} + k_{11} + k_{12} \\ .2\varepsilon k_{12}^2 + 10k_{12} \end{bmatrix}.$$

Thus, at each integration step the procedure requires solving the left hand side of (4) with respect to $dK/d\varepsilon$. This operation is simpler than solving (3) for a number of values of ε. In this example (3) is solved only for $\varepsilon = \varepsilon_o = 1/9$. This result is reported in detail in Proc. 3rd Princeton Conference on Systems and Information Theory, March 1969.

P. Kokotović
G. D'Ans
M. Jamshidi

13.4. Design of ε-Coupled Nonlinear Systems

Investigation was done on the applicability of decoupling in large optimization problems. The notion of ε-coupling developed for linear systems has been extended to nonlinear systems to achieve decoupling.

The results indicate that it is possible to obtain a near optimal solution for a large-scale optimization problem by successive solutions of low-order decoupled subsystems. The computational effort required for the near-optimal solution is considerably less than that required for the optimal one. Thus, the use of the near optimal solution can be justified.

The procedure involves:

(i) Introduction of a parameter ε into the system equations, as well as in the cost functional, such that at ε=0 the system decouples into smaller subsystems.

(ii) Expansion of the optimal control in a Taylor series about $\varepsilon=0$, truncated after a finite number of terms.

(iii) Solving for each of the terms in the series in (ii) to get the near-optimal control.

This procedure has been tested on two numerical examples. Both examples involved the motion of two particles under the influence of line charges. A satisfactory near optimal solution was obtained by using 3 terms of the expansion of the optimal control for most configurations.

It was also found that this procedure can provide a near optimal solution for problems with singularities in the accessory minimum problem. It could therefore overcome the difficulties encountered by using the second-variation methods for iterative solutions of such problems.

G. Singh
P. Kokotović

13.5. A Near-Optimal Controller for Nonlinear Systems

Any given nonlinear system-state equation can be represented by

$$\dot{X} = g(X,U) + \varepsilon f(X,U),$$

$$X(o,\varepsilon) = X^{o},$$

where X is n-dimensional state, U is m-dimensional control, g and f are linear and nonlinear functions of X and U, respectively, and ε is either a plant or an arbitrary parameter. Let g and the performance index J be the following,

$$g(X,U) = A(t)X + B(t)U + d(t)$$

$$J = \tfrac{1}{2}\int_{o}^{t_f} (X'Q_1X + U'R_1U)dt$$

where A(t) and B(t) are n×n and n×m matrices, d(t) is n-dimensional disturbance vector, and Q_1 and R_1 are n×n and m×m positive-definite matrices, respectively.

Two types of designs are considered, open loop and closed loop. The controller is an extension of "optimally sensitive" [1] design in the sense that an n^{th}-order truncated Taylor series expansion about $\varepsilon = 0$ (nominal system) is used. Control and state sensitivity equations are derived from the necessary conditions of the Maximum Principle. The

optimal control for the "nominal system" is of the linear state-regulator form which requires the solution of a matrix Riccati equation. The feedback block-transfer function of a closed-loop controller is the same as that of "nominal system" regardless of the number of terms considered in the series. The k^{th}-order sensitivity functions are obtained from the k-1^{st}-order and the nominal system trajectory and Riccati equation solutions.

The method was applied to a nonlinear plant with saturating amplifier and a 4th-order "single stand temper winding-mill plant." The latter example was an implementation of ARMCO Steel Corporation's No. 5 mill installation and its physical specification was used. The results indicate that with three terms in the Taylor series the near-optimal performance index is very close to the exact one. Further details may be found in CSL Report R-405, January 1969, by M. S. Jamshidi. The above controller was adapted for a third plant, "control logic for a regulator system" whose torque source is field-controlled d.c. motor and the result has been presented in a paper submitted to the Seventh Annual Allerton Conference on Circuit and System Theory, October 8-10, 1969. One other paper based on "Optimal Control Applications in Rolling Mills" is currently being prepared for publication.

M. Jamshidi

13.6. Singular Perturbation Method for Near Optimum Design of High-Order Nonlinear Systems

A direct application of a low-order design to a high-order plant may result in a performance far from optimum [2] or may even make the system unstable [4]. In [2], an approximation method has been proposed which represents a compromise between a simple, but unsatisfactory low-order design, and an optimal but computationally involved high-order design. The method proposed in [2] consists of approximating the high-order-optimal control $u^*(t,\lambda)$ by its truncated MacLaurin series in a scalar parameter λ,

$$\hat{u} = u^*(t,0) + \frac{\partial u^*}{\partial \lambda}(t,\lambda)\Big|_{\lambda=0} \lambda. \qquad (1)$$

The method in [2] is based on the assumption that the optimal control $u^*(t,\lambda)$ is continuous and differentiable with respect to λ at $\lambda = 0$. The perturbation

of λ from $\lambda > 0$ to $\lambda = 0$ is "singular" in the sense that at $\lambda = 0$ the order of the system is n while for $\lambda > 0$ the order is n+m.

This study reexamines the assumption made in [2] and proves, for a class of nonlinear systems, the continuity and differentiability properties of $u^*(t,\lambda)$ at $\lambda = 0$. An algorithm for calculating $\frac{\partial u^*}{\partial \lambda}(t,\lambda)$ at $\lambda = 0$ is also developed. All calculations are performed on a low-order (n-th order) model only, while approximation is valid for a high-order ((n+m)-th order) model.

In this study systems nonlinear in x, but linear in z and u are considered, that is

$$\dot{x} = g(x,z,u,\lambda,t) \equiv g_1 + B_1 z + C_1 u, \tag{2'}$$

$$\lambda \dot{z} \equiv G(x,z,u,\lambda,t) \equiv G_1 + B_2 z + C_2 u, \tag{2''}$$

where $g_1 = g_1(x,\lambda,t)$ and $G_1 = G_1(x,\lambda,t)$ are n and m dimensional vectors, and $B_1 = B_1(x,\lambda,t)$, $B_2 = B_2(x,\lambda,t)$, $C_1 = C_1(x,\lambda,t)$ and $C_2 = C_2(x,\lambda,t)$ are $n\times m$, $m\times m$, $n\times r$, and $m\times r$ dimensional matrices, respectively. It is assumed that g_1, B_1, and C_1 have continuous second partial derivatives and G_1, B_2, and C_2 have continuous third partial derivatives in a region Γ of the space of variables x, t, and with respect to λ for $\lambda \in [0,\lambda_{max}]$. It is also assumed that the performance index is quadratic in u,

$$J = \int_{t_o}^{T} V(x,u,t)dt = \frac{1}{2}\int_{t_o}^{T} [Q(x,t) + u'R(x,t)u]dt, \tag{3}$$

where $r\times r$ dimensional matrix $R(x,t)$ is positive definite in the region Γ and $R^{-1}(x,t)$ is assumed to have continuous second partial derivatives in the region Γ.

Let p and $\hat{q}$ be the adjoint variables corresponding to x and z, and let $q = \hat{q}/\lambda$. Then a necessary condition for the optimal control problem (2) and (3) is

$$\dot{x} = g_1 + B_1 z + C_1 u, \tag{4'}$$

$$\lambda \dot{z} = G_1 + B_2 z + C_2 u, \tag{4''}$$

$$\dot{p} = \nabla_x V - g_x' p - G_x' q, \tag{5'}$$

$$\lambda \dot{q} = -g_z' p - G_z' q, \tag{5''}$$

$$u = R^{-1}(C_1' p + C_2' q), \tag{6}$$

with the boundary values

$$x(t_o) = x_o; \quad z(t_o) = z_o; \quad p(T) = 0; \quad q(T) = 0. \tag{7}$$

Equations (4), (5), and (6) represent the high-order canonic equations. The low-order canonic equations, obtained by letting $\lambda = 0$, are

$$\dot{\underset{\sim}{x}} = \underset{\sim}{g}_1 + \underset{\sim}{B}_1\underset{\sim}{z} + \underset{\sim}{C}_1\underset{\sim}{R}^{-1}(\underset{\sim}{C}_1'\underset{\sim}{p} + \underset{\sim}{C}_2'\underset{\sim}{q}), \tag{8'}$$

$$0 = \underset{\sim}{G}_1 + \underset{\sim}{B}_2\underset{\sim}{z} + \underset{\sim}{C}_2\underset{\sim}{R}^{-1}(\underset{\sim}{C}_2'\underset{\sim}{q}), \tag{8''}$$

$$\dot{\underset{\sim}{p}} = \nabla_{\underset{\sim}{x}} V - \underset{\sim}{g}_{\underset{\sim}{x}}'\underset{\sim}{p} - \underset{\sim}{G}_{\underset{\sim}{x}}'\underset{\sim}{q}, \tag{9'}$$

$$0 = -\underset{\sim}{B}_1'\underset{\sim}{p} - \underset{\sim}{B}_2'\underset{\sim}{q}, \tag{9''}$$

$$\underset{\sim}{u} = \underset{\sim}{R}^{-1}(\underset{\sim}{C}_1'\underset{\sim}{p} + \underset{\sim}{C}_2'\underset{\sim}{q}), \tag{10}$$

with the boundary values

$$\underset{\sim}{x}(t_o) = x_o; \quad \underset{\sim}{p}(T) = 0. \tag{11}$$

The continuity and differentiability properties of the solution of the high-order boundary value problem are established by the following:

Theorem: If the solution $\underset{\sim}{x}^*(t)$ of the low-order equations (8) and (9) along with the condition (11) lies in the region Γ, and if the matrix $B_2(x,0,t)$ has all eigenvalues with negative real parts in the region Γ, then, as λ tends to zero, the high-order optimal functions $x^*(t,\lambda)$, $z^*(t,\lambda)$, $p^*(t,\lambda)$, $q^*(t,\lambda)$, and $u^*(t,\lambda)$ tend to the low-order optimal functions $\underset{\sim}{x}^*(t)$, $\underset{\sim}{z}^*(t)$, $\underset{\sim}{p}^*(t)$, $\underset{\sim}{q}^*(t)$ and $\underset{\sim}{u}^*(t)$ for all t, $t_o < t < T$. Furthermore, they are differentiable with respect to λ at $\lambda = 0$.

This theorem is proved by applying a result of Tupciev [3]. An important extension of the above theorem is possible when functions g and G are analytic in all of their arguments. It can be shown that in this case the functions $x^*(t,\lambda)$, $z^*(t,\lambda)$ $p^*(t,\lambda)$, $q^*(t,\lambda)$, and $u^*(t,\lambda)$ are analytic in λ at $\lambda = 0$, that is, that they can be expanded in an infinite MacLaurin series in λ.

It is shown now that the second term of the truncated series (1) can be obtained by solving a low-order matrix Riccati equation along with two low-order vector linear equations.

The introduction at $\lambda = 0$ of the sensitivity functions $\omega = \partial x^*/\partial\lambda$, $\sigma = \partial z^*/\partial\lambda$, $\pi = \partial p^*/\partial\lambda$, $\eta = \partial q^*/\partial\lambda$, $\nu = \partial u^*/\partial\lambda$, the differentiation of (4), (5), and (6) with respect to λ and letting $\lambda = 0$, are the steps by which

are low-order sensitivity equations are obtained,

$$\dot{\omega} = g_{\underset{\sim}{x}}\omega + \underset{\sim}{B}_1\sigma + \underset{\sim}{C}_1\nu + G_{\underset{\sim}{\lambda}}, \tag{12'}$$

$$0 = G_{\underset{\sim}{x}}\omega + \underset{\sim}{B}_2\sigma + \underset{\sim}{C}_2\nu + G_{\underset{\sim}{\lambda}} - \dot{z}^*, \tag{12''}$$

$$\dot{\pi} = -H_{\underset{\sim}{x}\underset{\sim}{x}}\omega - H_{\underset{\sim}{z}\underset{\sim}{x}}\sigma - H_{\underset{\sim}{p}\underset{\sim}{x}}\pi - H_{\underset{\sim}{q}\underset{\sim}{x}}\eta - H_{\underset{\sim}{u}\underset{\sim}{x}}\nu - H_{\underset{\sim}{\lambda}\underset{\sim}{x}}, \tag{13'}$$

$$0 = -H_{\underset{\sim}{x}\underset{\sim}{z}}\omega - H_{\underset{\sim}{p}\underset{\sim}{z}}\pi - H_{\underset{\sim}{q}\underset{\sim}{z}}\eta = H_{\underset{\sim}{\lambda}\underset{\sim}{z}} - \dot{q}^*, \tag{13''}$$

$$0 = H_{\underset{\sim}{x}\underset{\sim}{u}}\omega + H_{\underset{\sim}{p}\underset{\sim}{u}}\pi + H_{\underset{\sim}{q}\underset{\sim}{u}}\eta + H_{\underset{\sim}{u}\underset{\sim}{u}}\nu + H_{\underset{\sim}{\lambda}\underset{\sim}{u}}, \tag{14}$$

where H is the Hamiltonian, $H = \frac{1}{2}(Q + u'Ru) + p'(g_1+B_1z+C_1u) + q'(G_1+B_2z+C_2u)$.

The final value $\pi(T)$ is zero, while the initial value $\omega(t_o)$ is given by

$$\omega(t_o) = \int_{t_o}^{\infty} B_1(x_o,0,t_o)(z(\tau) - \varphi(x_o,\underset{\sim}{u}^*(t_o),t_o))d\tau \tag{15}$$

where $\varphi(x_o,\underset{\sim}{u}^*(t_o),t_o) = -B_2^{-1}(x_o,0,t_o)(G_1(x_o,0,t_o) + C_2(x_o,0,t_o)\underset{\sim}{u}^*(t_o))$, and the variable $z(\tau)$ is the solution of the equation $dz/d\tau = G_1(x_o,0,t_o) + C_2(x_o,0,t_o) + B_2(x_o,0,t_o)z(\tau)$ with the initial condition $z(\tau)\big|_{\tau=t_o} = z_o$.

With

$$\pi = -M\omega + \dot{\gamma},$$

this problem is reducible to a final-value problem for the n×n-dimensional, symmetric-matrix, Riccati equation

$$\dot{M} = -MA_1 - A_1'M+MA_2M - A_3; \quad M(T) = 0, \tag{16}$$

and an n-dimensional linear vector equation

$$\dot{\gamma} = (MA_2 = A_1')\gamma + M\delta_1 + \delta_2; \quad \gamma(T) = 0. \tag{17}$$

Then, ω can be obtained as the solution of the equation

$$\dot{\omega} = (A_1 - A_2M)\omega + A_2\gamma + \delta_1, \tag{18}$$

for the initial condition given by (15), and the control sensitivity ν can be evaluated from

$$\nu = \underset{\sim}{R}^{-1}((D_2-D_1M)\omega + D_1\gamma + H_{\underset{\sim}{\lambda}\underset{\sim}{u}} - H_{\underset{\sim}{q}\underset{\sim}{u}}(B_2')^{-1}(H_{\underset{\sim}{\lambda}\underset{\sim}{z}} + \dot{\underset{\sim}{q}}^*)), \tag{19}$$

where the expressions for A_1, A_2, A_3, δ_1, δ_2, D_1, and D_2 are easily obtained from (12), (13), and (14).

Hence, the obtaining of $\partial u^*/\partial\lambda$ as a function of t,λ at $\lambda = 0$ consists of the following steps:

Evaluation of the integral (15).

Solving of nXn-dimensional symmetric Riccati equation (16).

Solving of n-dimensional linear equations (17) and (18).

Since the evaluation of the integral (15) is straightforward, it can be seen from (16), (17), and (18) that the above singular perturbation method requires for an (n+m)th order plant essentially the same amount of computation as the neighboring optimum design for an n-th order plant.

The above result by Sannuti and Kokotović has been presented at the Fourth Congress of the IFAC, June 1969, Warsaw, Poland, and at the JACC, Boulder, Colorado, August 1969, and will be published in the IFAC journal Automatica, November 1969.

Further research in this area involves attempts to extend the above theorem to a broader class of systems and to incorporate the singular perturbation method in a general procedure for solving large-scale dynamic problems.

P. Kokotović
P. Sannuti
C. Hadlock

13.7. An Algorithm for Minimax Design

A minimax design of dynamic systems containing variable parameters has been considered. The variable parameters fall under two categories: those parameters represented by the vector x whose values can be selected by the designer, and those parameters, represented by the vector y, which can neither be preset nor precisely measured. A minimax design consists of selecting an x such that the maximum value of a scalar performance index J(x,y) with respect to y is minimized. The variables x* and y* are called minimax solutions if

$$J(x^*,y^*) = \min_{x\in X} \max_{y\in Y} J(x,y)$$

An algorithm has been developed for generating a convergent sequence of pairs $\{x_k, y_k\}$ whose limit satisfies necessary conditions for a minimax point. The following are assumptions that have been made in deriving the algorithm.

1) The admissible region X of x is in E_r and X is a closed convex bounded set.

2) The admissible region Y of y is in E_s and Y is a closed and bounded set.

3) J(x,y) and the partial derivative of J(x,y) with respect to x are jointly continuous with respect to both x and y for all admissible x and y except at isolated points in Y. At these isolated points, J(x,y) and the partial derivative of J(x,y) with respect to x are continuous with respect to x.

4) The function J(x,y) for any fixed admissible x has only a finite number of local maxima in Y. A local maximum is either a point $y_i \in Y$ whose corresponding value of J(x,y) is equal to or greater than the value of J for all admissible points of Y in a neighborhood of y_i, or an isolated point.

The algorithm is equally applicable for solving minimax problems whose solutions are saddle points, and for solving minimax problems whose solutions are not saddle points. Several numerical examples for which the solutions are not saddle points have been solved.

A Ph.D. thesis based on this work was submitted by J. E. Heller which also appears as CSL Report R-406. A paper has also been submitted for possible journal publication.

J. E. Heller
J. B. Cruz, Jr.
J. Medanić

13.8. Elimination-Type Minimax Algorithm

An elimination-type algorithm has been constructed for the solution of the following minimax problem: Find (x*,y*) such that

$$J(x^*,y^*) = \min_{x \in X} \max_{y \in Y} J(x,y)$$

where

(a) X, Y are compact and convex subsets of finite dimensional spaces,

(b) J(x,y) and $\partial J^*x,y)/\partial x$ are continuous jointly in their variables, and

(c) $J(x,y)$ is convex in X for all $y \in Y$.

Development of the algorithm and its application to minimax problems satisfying (a)-(c) has shown that it possesses the following desirable properties: (i) Maximization over Y, which is a mandatory task per iteration in many other proposed algorithms, need not be performed in each elimination in this algorithm; (ii) if the answering set $Y_o(x)$ is not a singleton, greater reduction per elimination is possible, and, hence, this does not reduce, but, in fact, increases the over all efficiency of the algorithm; (iii) the solution time is decreased with a good initial guess x_o for the minimax value x^*, because this usually demands a smaller number of complete maximizations as discussed in (i); (iv) because the elimination algorithm reduces the admissible domain to an arbitrary neighborhood of the minimax solution, the problem of an appropriate stopping criterion is inherently solved.

The algorithm is not restricted, as presently described, to problems where X is defined by linear constraints; nonlinear boundaries of X can be represented by a suitable number of supporting hyperplanes.

The major drawback of the algorithm is that it is restricted to functions $J(x,y)$ which are convex in X. Although it is possible to reduce somewhat this requirement, it nevertheless will restrict the generality of the algorithms. The project is described in Report R-421, "On Some Theoretical and Computational Aspects of the Minimax Problem," and a paper has been submitted for publication.

J. Medanić

13.9. Feedback Strategies for a Differential Game with Singular Solutions

A sixth-order bilinear differential game with three control variables, based on the physical problem of controlling the motion of a charged particle by magnetic fields, has been studied. The system is described by

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = u_3 x_4 - v x_6,$$

$$\dot{x}_3 = x_4,$$

$$\dot{x}_4 = u_1 x_6 - u_3 x_2,$$

$$\dot{x}_5 = x_6,$$

$$\dot{x}_6 = -u_1 x_4 + v x_2.$$

The controls are amplitude bounded,

$$|u_i| \leq k_i,$$

and the terminal manifold is

$$x_1^2 + x_3^2 + x_5^2 = \delta^2.$$

The payoff is the time of termination. The primary aim was to obtain from the necessary conditions the candidates for optimal strategies, with the ultimate aim of synthesizing the complete feedback solution from the obtained candidates. The problem admits singular solutions, and the necessary conditions for optimality in singular regimes were used to extract those singular regimes (determined by the rank and degree of singularity) which may be optimal. It was shown that, in a region neighboring the terminal manifold feedback strategies,

$$u_1 = k_1 \text{sign}(x_4 x_5 - x_3 x_6),$$

$$v = -k_2 \text{sign}(x_1 x_6 - x_2 x_5),$$

$$u_3 = k_3 \text{sign}(x_2 x_3 - x_1 x_5),$$

satisfy the necessary conditions for optimality in both the normal and various singular regimes. A complete synthesis however, has not yet been achieved because of the abundance of singular solutions for the players. The preliminary results are described in the paper "Feedback Strategies for a Differential Game with Singular Solutions," to be presented at the First International Conference on Differential Games, Amhurst, Mass., 1969.

J. Medanić
C. I. Chen

13.10. Suboptimal Filtering and Control of Stochastic Systems

The Kalman-Bucy filter generates the best mean-square estimate, $\hat{x}(t,\omega)$, of a Gauss-Markov random process described by a linear stochastic differential equation of the Ito type:

$$d\underset{\sim}{x}(t,\omega) = [A\underset{\sim}{x}(t,\omega) + B\underset{\sim}{u}(t,\omega)]dt + Gd\underset{\sim}{B}(t,\omega),$$
$$\underset{\sim}{x}(t_o) = \underset{\sim}{x}_o, \qquad t\in(t_o,t_f) . \tag{1}$$

In this, $\underset{\sim}{x}_o$ is a zero-mean, Gaussian, random variable with covariance $P_o = E\{\underset{\sim}{x}_o \underset{\sim}{x}_o'\}$; $\underset{\sim}{B}_t$ is a Brownian-motion process which is uncorrelated with x_τ, $\tau > t$ and $E\{d\underset{\sim}{B}_t\, d\underset{\sim}{B}_t'\} = Q(t)dt + \theta(dt)$. The observations or measurements available to the filter are linear in $\underset{\sim}{x}(t,\omega)$:

$$d\underset{\sim}{Z}(t,\omega) = H\underset{\sim}{X}(t,\omega)dt + d\underset{\sim}{V}(t,\omega). \tag{2}$$

In this, $\underset{\sim}{V}_t$ is a Brownian motion process with $E\{d\underset{\sim}{V}_t\, d\underset{\sim}{V}_t'\} = R(t)dt + \theta(dt)$; let F_t denote the minimum sigma field generated by the data $A_{t_o}^{\,t} = \{Z_s \,|\, s\in(t_o,t)\}$. The matrices A, B, G, and H may be time-varying and are assumed to be a function of a scalar parameter q.

The estimate is generated by the following linear stochastic system driven by the observations:

$$d\hat{\underset{\sim}{x}}(t,\omega) = [A\hat{\underset{\sim}{x}}(t,\omega) + B\underset{\sim}{u}(t,\omega)]dt + K[dZ_t - H\hat{\underset{\sim}{x}}(t,\omega)dt],$$
$$\hat{\underset{\sim}{x}}(0,\omega) = \underset{\sim}{0}, \tag{3}$$

where

$$K = P_t H' R^{-1} ,$$
$$P_t = E\{(\underset{\sim}{x}-\underset{\sim}{x})(\underset{\sim}{x}-\underset{\sim}{x})'\}$$

and P_t satisfies the nonlinear differential equation of the Riccati type,

$$\dot{P} = AP + PA - PH'R^{-1}HP,$$
$$P(t_o) = P_o . \tag{4}$$

The performance of the filter, the mean square error of estimation, is given by the trace of the covariance of error matrix:

$$J_t = E\{\|\underset{\sim}{x}-\hat{\underset{\sim}{x}}_t\|^2\} = tr\{P_t\}. \tag{5}$$

Nonoptimal estimates will be denoted by $\hat{\hat{\underset{\sim}{x}}}(t)$ and the resulting estimation error will be denoted by

where

$$\tilde{J}_t = E\{\|\underset{\sim}{x}_t-\hat{\hat{\underset{\sim}{x}}}\|^2\} = tr\{\tilde{P}_t\},$$
$$\tilde{P}_t = E\{(\underset{\sim}{x}_t-\hat{\hat{\underset{\sim}{x}}}_t)(\underset{\sim}{x}_t-\hat{\hat{\underset{\sim}{x}}}_t)'\}. \tag{6}$$

Equations have been derived which describe the degradation in performance of the filter when the matrices A, B, G, H, R, and Q are not known or implemented exactly into the filtering system (3). As can be seen there are many combinations, some of which are special cases of others.

These equations will not be presented here. These equations can be used in designs to minimize the effect of parameter variations. For the case when the system matrices are analytic in the parameter q and modeling implementation errors are described by truncations in Taylor series, more exact descriptions of the performance degradation are available. Two of the more interesting results are presented here:

A) A truncated gain

$$\tilde{K} = \sum_{i=o}^{n} \frac{\partial^i K}{\partial q^i} \frac{q^i}{i!}$$

is used in the filter instead of the optimal gain matrix G: the filter is described by

$$d\hat{\hat{x}}_t = [A\hat{\hat{x}}_t + Bu]dt + \tilde{K}[dz_t - H\hat{\hat{x}}_t dt] \tag{7}$$

and its performance by $\tilde{J}_t$ defined in (6) above. Then $\tilde{J}_t$ matches J_t in q to the (2n+1) order;

$$\left.\frac{\partial^i \tilde{J}_t}{\partial q^i}\right|_{q=0} = \left.\frac{\partial^i J_t}{\partial q^i}\right|_{q=0} \qquad i = 0,1,\ldots,(2n+1). \tag{8}$$

B) Instead of the ideal filter, described by (3), the following filter is used:

$$d\hat{\hat{x}}_t = [\tilde{A}\hat{\hat{x}}_t + \tilde{B}u_t]dt + \tilde{P}_t\tilde{H}R^{-1}[dz_t - \tilde{H}_t\hat{\hat{x}}_t dt], \tag{9}$$

where

$$\tilde{A} = \sum_{i=o}^{\alpha} \frac{\partial^i A}{\partial q^i} \frac{q^i}{i!}, \quad \tilde{B} = \sum_{i=o}^{\beta} \frac{\partial^i B}{\partial a^i} \frac{q^i}{i!}$$

$$\tilde{G} = \sum_{i=o}^{\gamma} \frac{\partial^i G}{\partial q^i} \frac{q^i}{i!}, \quad \tilde{H} = \sum_{i=o}^{\lambda} \frac{\partial^i H}{\partial q^i} \frac{q^i}{i!}.$$

The performance of this filter, $\tilde{J}_t$, matches that of the optimal filter to (2n+1) order again (in the sense of (8)) where

$$n = \min\{\alpha, \beta, \gamma, \lambda\}. \tag{10}$$

The above results all have valid discrete time versions.

Consider the stochastic optimal-control problem of finding a feedback control $u(t,\omega)$ which minimizes the following quadratic performance index for each t:

$$V(\hat{x},t) = \tfrac{1}{2}\, E\{\|\underset{\sim}{x}(t_f)\|^2_{Q_f} + \int_t^{t_f} [\|\ \|^2_{Q_o} + \|\underset{\sim}{u}_s\|^2_{R_o}]ds \,|\, F_t\} \tag{11}$$

For the linear plant described by (1), the solution is well known to be

$$\begin{aligned} u_t &= -M(t)\hat{\underset{\sim}{x}}(t) \\ M(t) &= R_o^{-1} B' \Sigma(t) \end{aligned} \tag{12}$$

where $\Sigma(t)$ satisfies a nonlinear differential equation of the Riccati type

$$-\dot{\Sigma}_t = A'\,\Sigma_t + \Sigma_t A - \Sigma_t B R_o^{-1} B \Sigma_t + Q_o$$

and $\hat{\underset{\sim}{x}}(t)$ is the best mean square estimate of $\underset{\sim}{x}(t)$ as generated by the Kalman filter, (3). This is just the separation principle for linear systems.

Analogous to the Cruz-Kokotović-Werner results for deterministic systems, it has been shown that if instead of the optimal gain matrix, $M(t)$, a truncated matrix $\tilde{M}(t)$,

$$\tilde{M}(t) = \sum_{i=o}^{m} \frac{\partial^i M}{\partial q^i} \frac{q^i}{i!} \tag{13}$$

is used, the performance $\tilde{V}(\hat{x},t)$ of the resulting system matches $V(\hat{x},t)$ to $(2m+1)$ in q:

$$\left.\frac{d^i V[\hat{x},t]}{dq^i}\right|_{q=0} = \left.\frac{d^i \tilde{V}[\hat{x},t]}{dq^i}\right|_{q=0} \quad ; \; i = 0,1,2,\ldots,(2m+1) \tag{14}$$

The results presented in the above paragraphs can be combined to form an error-separation theorem for linear stochastic systems:

Let the control $u(t,\omega)$ applied to the system described by (1) be given by

$$\tilde{u}(t,\omega) = -\tilde{M}(t)\hat{\hat{\underset{\sim}{x}}}(t)$$

where $\tilde{M}$ is given by (13) and $\hat{\hat{\underset{\sim}{x}}}(t)$ is an estimate of $\underset{\sim}{x}(t)$ such that

$$\left.\frac{d^i J_t}{dq^i}\right|_{q=0} = \left.\frac{d^i \tilde{J}_t}{dq^i}\right|_{q=0} \qquad i = 0,.,\ldots,2n+1$$

Then the performance, $\tilde{V}[\hat{\hat{x}},t]$, of the suboptimal system matches the optimal performance $V[\hat{x},t]$ to $(2p+1)$ in q where $p = \min\{n,m\}$; that is

$$\left.\frac{d^i V[\hat{x}_t,t]}{dq^i}\right|_{q=0} = \left.\frac{d^i \tilde{V}[\hat{\hat{x}}_t,t]}{dq^i}\right|_{q=0} \quad ; \; i = 0,1,\ldots,(2p+1).$$

The discrete time analog is also valid.

The above results assume that q is a known parameter. Some results are available when q is replaced by $\hat{q}$ in the above systems; of course, one still has to obtain $\hat{q}$ in some "optimum" fashion. In the ideal case, one would have an optimally-adaptive estimating scheme. At this point, it seems one should recall that the best mean-square estimate is defined by

$$\hat{x}(t) = E\{x(t)|F_t\} = \int \underset{\sim}{x}(t)\ \underset{\sim}{p}(x(t)|F_t)d\underset{\sim}{x}.$$

In general, $\underset{\sim}{p}(\underset{\sim}{x}(t)|F_t)$ is given by a functional equation containing stochastic integrals, which describes its evolution in time. For linear dynamics, this results in the Kalman filter. However, there are only a few other special cases which have been solved. The classical approach is to linearize around a nominal trajectory and then estimate the variation from the nominal trajectory. A different and promising approach is to expand the functional expression for $p(x|F_t)$ around $\hat{x}(t)$. The results are the same for linear systems with known parameters but quite different in even the slightly nonlinear cases investigated to this point.

M. Rice

13.11. Sensitivity of Distributed-Parameter Systems

Distributed-parameter systems arise naturally in many engineering problems involving phenomena such as vibration, heat propagation, diffusion, and fluid flow. The literature now contains many works concerning the optimal control of distributed-parameter [5] systems. To a lessor extent the questions of stability, controllability, and observability [6] have been considered. However, relatively little has been done concerning the sensitivity of distributed parameter systems. To date only two articles [7,8] have been concerned with this question.

These articles have attempted to reformulate sensitivity results for lumped parameter systems so that they may be used for distributed parameter systems. The main disadvantage of the resultant procedure is that the partial differential equation describing the system have to be separable and thus the method is very restricted and usually impractical.

The basic result of investigations to date has been the development of a sufficient condition to insure that feedback compensation

decreases the effects of parameter variations upon system dynamics as compared to an equivalent open-loop system subjected to the same parameter variations. The condition is easily tested and is valid for linear time-invariant distributed parameter systems which may be described by non-separable partial differential equations. The procedure is being applied to the process of heat exchange between a stationary and mobile medium.

J. Davis

13.12. Application of Optimal Control Theory to the Traffic Problem

Optimal control theory is being applied to the traffic control of an expressway system. An origination-destination matrix is assumed. In the first stage of design, the input rate is found by optimization technique, so as to minimize average time of waiting and passing. In the second state of design, the methods of kinematic wave propagation and sensitivity are being applied to maintain the optimality and prevent congestion.

C. I. Chen

13.13. Design of Model Following Systems Using the Companion Transformation

An approach used in many control systems designs is to force a controlled plant to respond approximately like a model [9-12]. That is, given a model previously selected to exhibit a desirable response to an input, a variable-parameter vector $\underset{\sim}{v}$ is adjusted in a controlled system so that the system and model output vector $\underset{\sim}{y}$ and $\underset{\sim}{y}^m$ are as close as possible. This design approach has been used for model-reference adaptive control, where the parameter vector is adjusted to minimize effects of variable plant parameters on the system response [9-11]. In this case, the model is often a model of the plant and controller with all parameters set to their nominal design values.

In this study, a new approach to the model following problem for single-input, linear, time-invariant systems is presented. Assume that the linear, time-invariant system to be controlled is described by the state equations

$$\dot{\underset{\sim}{x}} = A(\underset{\sim}{v})\underset{\sim}{x} + b(\underset{\sim}{v})u \tag{1}$$

$$\underset{\sim}{y} = C\underset{\sim}{x},$$

and the model is described by

$$\dot{\underset{\sim}{x}}^m = A_m\underset{\sim}{x}^m + \underset{\sim}{b}_m u_o$$
$$\underset{\sim}{y}^m = C_m\underset{\sim}{x}^m, \tag{2}$$

where

$\underset{\sim}{x},\underset{\sim}{x}^m$ are n dimensional state vectors

$\underset{\sim}{y},\underset{\sim}{y}^m$ are p dimensional output vectors

$\underset{\sim}{v}$ is an r dimensional parameter vector

u,u_o are scalar inputs.

It is desired to adjust $\underset{\sim}{v}$ to make $\underset{\sim}{y}$ and $\underset{\sim}{y}^m$ as close as possible in some sense. The approach here is to minimize a bound on the norm $\|\underset{\sim}{y}-\underset{\sim}{y}^m\|$. This is shown to result in an optimization involving only algebraic equations (as opposed to differential equations). Moreover, examples show this approach can lead to a fast numerical minimization procedure yielding results which are very close to those obtained by actually minimizing $\|y-y^m\|$.

Since experience has suggested that a special choice of state variables may lead to an easier numerical minimization procedure, consider two nonsingular, time-invariant, linear transformations T and T_m of the system and model state vectors, and denote the transformed states by $\underset{\sim}{z}$ and $\underset{\sim}{z}^m$, so that

$$\underset{\sim}{x} = T(\underset{\sim}{v})\underset{\sim}{z}, \quad \underset{\sim}{x}^m = T_m\underset{\sim}{z}^m. \tag{3}$$

Note that in general T is allowed to depend on $\underset{\sim}{v}$.

In terms of the transformed state variables, the problem of causing $\underset{\sim}{y}$ to follow $\underset{\sim}{y}^m$ by minimizing the norm of their difference becomes

$$\min_{\underset{\sim}{v}}\|C\underset{\sim}{x}-C_m\underset{\sim}{x}^m\| = \min_{\underset{\sim}{v}}\|(CT(\underset{\sim}{v})\underset{\sim}{z} - C_mT_m)\underset{\sim}{z}^m\|. \tag{4}$$

It would be desirable if the eigenvalues of the system could be forced to be the same as the model, since these characterize the zero-input response of the system. The system and model may then be called "dynamically similar" to the model independent of the choice of the state coordinates.

Pursuing this approach, it becomes apparent that a particularly suitable transformation to use in (3) and (4) is the transformation to companion form, i.e., the A matrix in the state description of the transformed system is in companion form [13,14].

It is shown that forcing the system to have the same eigenvalues as the model implies that $\underset{\sim}{z} = \underset{\sim}{z}^m$. The problem in (4) becomes

$$\min_{\underset{\sim}{v}} \|\underset{\sim}{y}-\underset{\sim}{y}^m\| = \min_{\underset{\sim}{v}} \|(CT(\underset{\sim}{v})-C_mT_m)\ \underset{\sim}{z}^m\| \leq \min\{\|CT(\underset{\sim}{v})-C_mT_m\|\ \|\underset{\sim}{z}^m\|\} \tag{5}$$

where the operator norm indicated in (5) is a matrix norm. It is shown that by choosing to minimize the upper bound

$$\min_{\underset{\sim}{v}} \|CT(\underset{\sim}{v}) - C_mT_m\| \tag{6}$$

on $\|y-y^m\|$, the considered minimization problem is an algebraic optimization. This minimization indicated in (6) is logical, since if $\underset{\sim}{z} = \underset{\sim}{z}^m$, then in order for $\underset{\sim}{y}$ to be close to $\underset{\sim}{y}^m$ the difference of the transformations relating $\underset{\sim}{y}$ and $\underset{\sim}{y}^m$ to $\underset{\sim}{z}$ and $\underset{\sim}{z}^m$ should be small.

The advantages of this technique over those requiring minimization of an integral square error between the system and model outputs [9-11] are that no system simulations are required, no averaging time for the integration is required, the result is independent of the system input as a time function, and the system and model need not be repeatedly excited in real-time optimization as in some model reference-adaptive schemes.

W. R. Perkins

13.14. References

1. Kokotović, P. V., J. B. Cruz, Jr., J. E. Heller, and P. Sannuti, "Synthesis of Optimally Sensitive Systems," Proceedings of the IEEE, 56, 1318-1324, (August 1968).

2. Kokotović, P. and P. Sannuti, "Singular Perturbation Method for Reducing the Model Order in Optimal Control Design," Preprints of Joint Automatic Control Conference, June, 1968. See also IEEE Transactions on Automatic Control, 13, August, 1968.

3. V. A. Tupciev, "Asymptotic Behavior of the Solution of a Boundary-Value Problem for Systems of Differential Equations of First Order with a Small Parameter in the Derivative," Dokl. Akad. Nauk SSSR 143, (1962), 1296-1299. See also: V. A. Tupciev, "On 'Angular' Solutions of Boundary Value Problems with a Small Parameter Multiplying a Derivative in a System of First Order Equations," Vstn. Mosk, Univ. Ser. I, Mat. Meh., No. 3, (1963), 17-24.

4. P. Sannuti and P. Kokotović, "An Approximate Design of Optimal Regulators for High-Order Linear Plants," Proceedings of the Second IFAC Symposium on System Sensitivity and Adaptivity, Dubrovnik, Yugoslavia, August 26-31, 1968.

5. W. L. Brogan, "Optimal Control Theory Applied to Systems Described by Partial Differential Equations," Advances in Control Systems 6, Chapter 4 (1968).

6. P. K. Wang, "Control of Distributed Parameter Systems," Advances in Control Systems 1, Chapter 3 (1964).

7. W. A. Porter, "Sensitivity Problems in Distributive Systems," Int. J. Control, 5, 393, (1967).

8. W. A. Porter, "Parameter Sensitivity in Distributed Feedback Systems," Int. J. Control, 5, 413 (1967).

9. H. P. Whitaker, "Design Capabilities of Model Reference Adaptive Systems," Proc. NEC, Vol. XVIII, 1962, pp. 241-249.

10. P. C. Parks, "Liapunov Redesign of Model Reference Adaptive Control Systems," IEEE Trans. on Automatic Control, Vol. AC-1, No. 3, July 1966, pp. 362-367.

11. P. V. Kokotović, et. al, "Sensitivity Method in the Experimental Design of Adaptive Control Systems," Proc. Third IFAC Congress, London, June 1966, pp. 45B.1-45B.12.

12. T. C. Coffey, "The Application of Modern Computing Technology to Control System Analysis and Design Problems," Defense Documentation Center Document AD660-133, June 1967, pp. 63-107.

13. W. M. Wonham and C. D. Johnson, "Optimal Bang-Bang Control with Quadratic Performance Index," Preprints Fourth JACC, 1963, pp. 102-112.

14. D. S. Rane, "A Simplified Transformation to (Phase-Variable) Canonical Form," IEEE Trans. on Automatic Control, Vol. AC-11, No. 3, July 1966, p. 608.

14. SWITCHING SYSTEMS†

G. Metze
H. Finkelstein
T. Gaddess
J. Hayes
M. Paige
T. Powell
D. Schertz
S. Seth
R. Shrader
A. Wojcik

14.1 Diagnosability Improvement by Test-Point Insertion

To improve module-level diagnosability of modular, combinational, logic circuits, three techniques for allocating test points have been developed. The first, analogous to the prime-implicant covering problem of classical switching theory, requires a large amount of computation but produces a minimum test-point set. The second, a graphical worst-case analysis, is considerably more economical but produces nonoptimum results. The third is a combination of the first two and lies between them in computation required and optimality of results.

It is proved that the graphical test-point specification algorithm is sufficient to guarantee complete module-level diagnosability of the circuit if the test points are used to gain more information on the result of each test. A general graph-theoretic characterization of the nonoptimum set of selected test points is given in terms of accessibility properties. A final theorem is proved to show that tests derived on the basis of fault detection only are sufficient for diagnosing the previously troublesome fault classes.

These results are presented in Report R-409[1] together with an efficient notation useful for analyzing the behavior of a combinational circuit under all allowed fault conditions. The procedure is algorithmic in nature and programmable, and flow charts are given to illustrate one possible implementation.

T. Gaddess

14.2 Fault Diagnosis of Combinational Logic Modules

A method to diagnose modules of combinational logic has been developed. Included is a procedure to generate a set of detection tests for a module, called the Fault Detection procedure (FD-procedure), and a method for generating a set of diagnostic tests for a combinational circuit of modules. The method for

†This work was supported in part by the National Science Foundation under Grant NSF-GK-1663 and the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract No. DAAB-07-67-C-0199.

generating a diagnostic set of tests first finds a set of tests which detects the faults in the total circuit, determines which faults in different modules need to be identified, and generates the remaining necessary tests by an application of an exclusive-or operation to complete the diagnosis.

A proof that the FD-procedure generates a set of tests which detects all detectable faults for a module is given. The method to generate diagnostic tests always identifies every faulty module that can be identified because of the exclusive-or operation. Examples and comparisons to other approaches are given.

Results of this study are presented in Report R-413[2].

T. Powell

14.3. Single and Multiple Faults in Combinational Networks

A new representation for faults in combinational digital systems has been developed. Faults which are inherently indistinguishable are identified and combined into classes, and the behavior of the circuit under fault conditions is represented in terms of these classes. This results in a description of the faulty circuit by means of Boolean equations that are readily manipulated for the purpose of test generation or fault simulation. A connection-graph interpretation of this fault representation can be used to derive heuristic methods for the selection of efficient tests without extensive computation.

The fault classes form a geometric structure which effectively subdivides the original circuit into fanout-free segments. This fanout-free characteristic allows a simplified analysis of multiple fault conditions. In particular, it can be used to show that, given any set of tests that is guaranteed to detect all single faults, there exists a large class of networks for which all multiple faults will be detected by the complete single-fault test set, and, for networks not belonging to this class, the multiple faults that are not guaranteed to be detected by the complete single-fault test set can be identified, and multiple-fault tests can be easily generated for them. This work is described in Report R-418[3].

D. R. Schertz

14.4. The Design of Diagnosable Sequential Machines

Work was continued on the investigation of the fault-diagnosis properties of digital networks. In particular, certain classes of networks which are optimal from the diagnosis point of view, i.e., networks with test sets that are easily derived and minimal in some sense, have been found, and their properties examined.

Some unexpected limits on the minimum number of tests needed for any realization of a Boolean function have been obtained. These results are currently being prepared for publication.

J. P. Hayes

14.5 Fault Diagnosis in Regular Arrays

The work on the diagnosis of two-dimensional, regular, combinational arrays continues. The conditions under which all input combinations are applicable to any cell in an arbitrary-size array were investigated. This condition is isomorphic to the existence of tessellations (finite or infinite) for every input combination. For diamond-shape, finite tessellations of size 2Xn, n arbitrary, it is proved that such tessellations do not exist for (1) the constant functions "zero" and "one," and (2) the logically incomplete nodal functions of two variables: AND, IMP, NIMP, and their duals. A nonexhaustive algorithm to find larger-sized, diamond-shape tessellations has also been found. Currently the conditions of propagation of a fault from its origin to the observable output boundaries of an array are being investigated.

S. Seth

14.6. N-Valued Switching Theory

A paper describing the logic development of ternary switching theory, and discussing some of the problems that arise both in the logical development and physical implementation of combinational and sequential ternary circuits, has been concluded[4]. A discussion of both combinational and sequential ternary switching theory is presented, and some of the similarities to and differences from binary switching theory are indicated. The topics of hazards, variable-value ordering schemes, and ternary memory devices are examined and

pointed out to be areas that need further development.

Currently, the study of ternary and n-valued logics is continuing. Preliminary results on the use of n-valued memory devices have been obtained. Minimization of multioutput, n-valued, combinational circuits is being studied in conjunction with relative cost analysis of n-valued adder networks. The main emphasis has been on the formulation of n-valued sequential switching theory.

A. Wojcik

14.7. Generation of Diagnostic Tests using Prime Implicants

The question of whether it is possible to derive a reasonable set of fault tests with a minimum of analysis of the structural details of a logic network has been investigated. It has been found that consideration of the prime implicant information and knowledge of whether the network is realized in the sum-of-products or the product-of-sums form (or their NAND/NOR equivalents) leads to a sufficient set of fault detection tests, even for multilevel networks. This work is described in Report R-414[5].

M. Paige

14.8. References

1. T. G. Gaddess, Improving the Diagnosability of Modular Combinational Logic by Test-Point Insertion, Technical Report R-409, Coordinated Science Laboratory, University of Illinois (March 1969).

2. T. S. Powell, A Module-Diagnostic Procedure for Combinational Logic, Technical Report R-413, Coordinated Science Laboratory, University of Illinois (April, 1969).

3. D. R. Schertz, On the Representation of Digital Faults, Technical Report R-418, Coordinated Science Laboratory, University of Illinois (May, 1969).

4. A. S. Wojcik and G. Metze, "A Reexamination of Ternary Switching Theory," Proc. Sixth Allerton Conference on Circuit and System Theory, pp. 711-721, October, 1968.

5. M. R. Paige, Generation of Diagnostic Tests Using Prime Implicants, Technical Report R-414, Coordinated Science Laboratory, University of Illinois (May, 1969).

H. Von Foerster
R. T. Chien
P. Arnold
K. Biss
B. Carss
J. Chow
P. Duran
J. Lipovski
E. Mark
F. P. Preparata
S. Ray
T. Woo
J. Schultz
F. Stahl
S. Taylor
P. Weston
M. Wilkins
V. Wu

15.1. Research on the R2 System

15.1.1. General Remarks

Research on the R2 (Rules of the Road) system has progressed to an advanced stage. As is outlined in the architecture, natural-language questions will be converted into formal-language questions with a semantically directed parser. Work is continuing on semantic modeling, parser implementation and formal-language conversions. Plans are being carried out to install a LISP 1.5 processor. This new facility should be very useful to us.

15.1.2. Architecture of the R2 System

An overall architectural approach to the R2 question-answering system has been undertaken. Figure 15.1 contains a graphical representation of the most recent scheme.

The R2 system will be constructed in a modular fashion with a mechanism that passes CONTROL from one module to another. The justification for modularization is twofold:

a. A variety of different organizations of the modules can be accomodated with a minimal amount of overall system modification. For example, suppose it were required that the modules were used in the order A,B,C,D,... on one occasion and in the order A,D,B,C,... at some other time.
b. New modules can be added without disrupting previous module organizations. That is, when a new module is added it can be used with the other modules at the will of the system designer.

There is a section of the CONTROL called the EXECUTOR. The EXECUTOR is simply a method of allowing a program created by the system to be executed and to have control pass back to the CONTROL unit upon termination of the program.

The MEMORY is intended to be available to each module and in addition to the CONTROL. This is not to say that memory used by one module

†This work was supported by the Office of Education under Contract No. OE C-1-7-071213-4557 and by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB-07-67-C-0199.

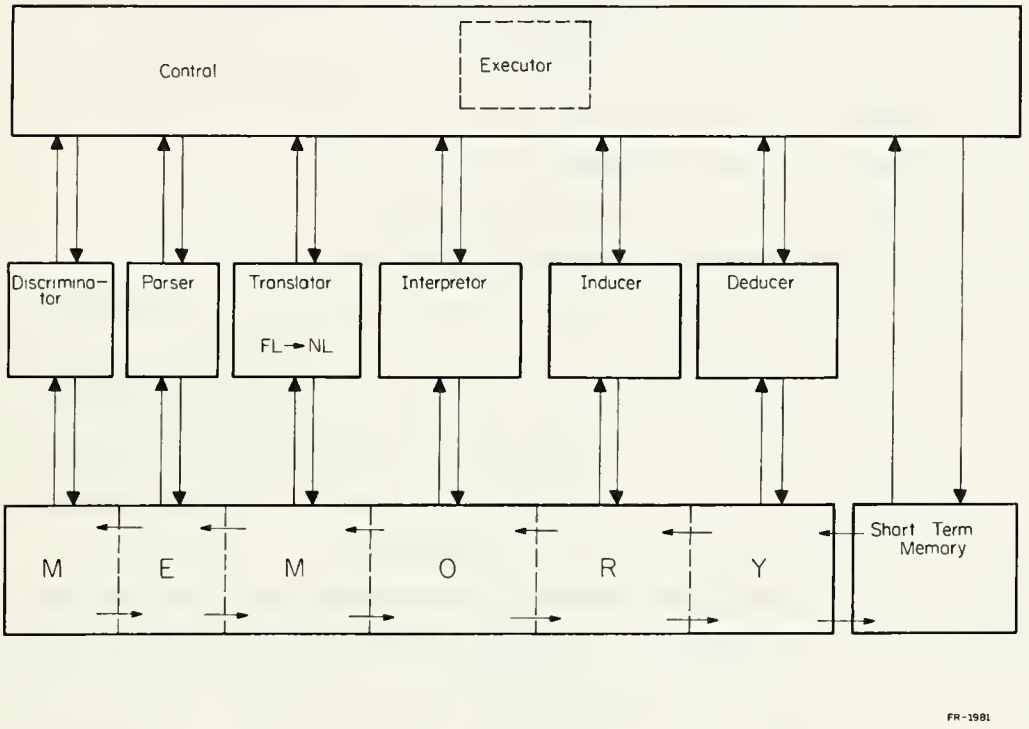


Fig. 15.1. Architecture of the R-2 system.

cannot be used by another module. Short-term memory is indicated and served as a scratch area for the CONTROL to store input and output from the various modules in addition to storing intermediary results.

A partial description of the structural organization of the memory is indicated in the description of some of the modules. It should be noted, however, that a version of LISP 1.5 written for IBM system 360 is to be used for some if not all structural organization.

The following are descriptions of the individual modules:

a. PARSER

input--a natural language (NL) string

output--a formal language (FL) string

The rules governing formal language strings are:

1. Verbs and function words are treated as n-place functions.

   e.g., Cars have wheels. (have(cars,wheels))

   An octagon sign means stop. (means(octagon sign,stop))

   Cars and trucks have wheels. have(and(cars,trucks),wheels)

2. Modifiers are treated as functions with the structure being modified appearing as an argument.

   e.g., Cars have wheels. (have(plural(car),plural(wheel))

   An octagon sign means stop. mean(octagon(sign),stop)

   If a driver has a valid driver's license, he may drive a car.

   If-then(have(a(driver),a(valid(driver's(license)))),

   drive(a(driver),a(car))

For a more detailed description of the parser see Sec. 15.1.3.

b. DISCRIMINATOR

input--an FL string

output--a set of contextual disagreements in usage

The DISCRIMINATOR compares the usage of the elements of the parsing with previous usages and sets up a dialogue with the PARSER on the parsing that best fits the word usages or is a match. This technique has been adapted from Siklossy [1].

c. INTERPRETOR

input--an FL string

output--a set of parameters or a program

It is assumed that ambiguities have been resolved to some extent before reaching this module. Note, however, that the interpretation in the

sense of the parsing may still be incorrect. The first function of the interpretor to determine if the FL string is a statement of fact, a command, or a question.

1. If the FL string is a statement of fact, all necessary information in order to insert this fact into the memory is extracted, and a program is constructed for the executor to inset the information.
2. If the FL string is a command, the necessary instructions are generated, and the executor carries out the command.
3. If the FL string is a question, then the necessary information is extracted, and a program is given to the executor in order to retrieve some information from the memory.

Some of the techniques used by the interpretor have been adapted from Woods [2].

d. DEDUCER

input--an FL string from the user
an FL string from memory

output--an FL string

This module is intended to work with the first-order predicate calculus to prove or disprove the answer to the question. For background material, see Simmons [3], Schwarcz [4] and Robinson [5,6].

e. INDUCER

This is a very sketchy module intended to make some good guesses as to what the next move should be if the deducer has disproved the answer. Possibilities include finding alternative parsings, fetching other FL strings from the memory, or asking the user to clarify certain aspects of the question.

f. TRANSLATOR

input--an FL string

output--an NL string

An FL string is converted to an NL equivalent. This module will not be incorporated in the first implementation of R2.

Fred Stahl

15.1.3. Description of the Parser (Syntactical Analysis of the R-2 System)

I. Given Information to the parser

A. A dictionary of the words of the data base. Each word will have its 3 parts of speech (or less) with its most-common use first, second-most-common use next, and its least-common use third.

B. A set of Immediate-Constituents rules with order.

C. A set of transformational rules with ordering.

D. A set of phrase markers defined in a mini-dictionary.

II. The Main Program

A. A natural language string defined to be a grammatically correct sentence or question S.

B. Algorithm

1. Check to see if it is an if-then type of sentence
2. Check for coordinating conjunctions
3. Transform to a complex sentence S'
   a. Check for relative pronouns
   b. Check for subordinating conjunctions
   c. Check for conjunctives
   d. Check for participle phrases
   e. Check for gerundive phrases
   f. Check for the number of verbs in S'
4. Transform to simple sentence S''.
   a. Check for infinitives
   b. Use Immediate Contituency Analysis to parse the rest of S (which is now S'') to find out what is the main verb, subject, predicate nominitive and/or object.

C. Output:

The form of the output will be:

Main Verb (modifiers)(subject(modifiers), object(modifiers))

Substitutions and its structure: (Example: R = relative pronoun (verb(modifiers)) (subject(modifiers), object (modifiers)))

III. Description of the Program (See Fig. 15.2)

B-1. Check for if then sentence

This is a simple test to see if one of four conditions exist:

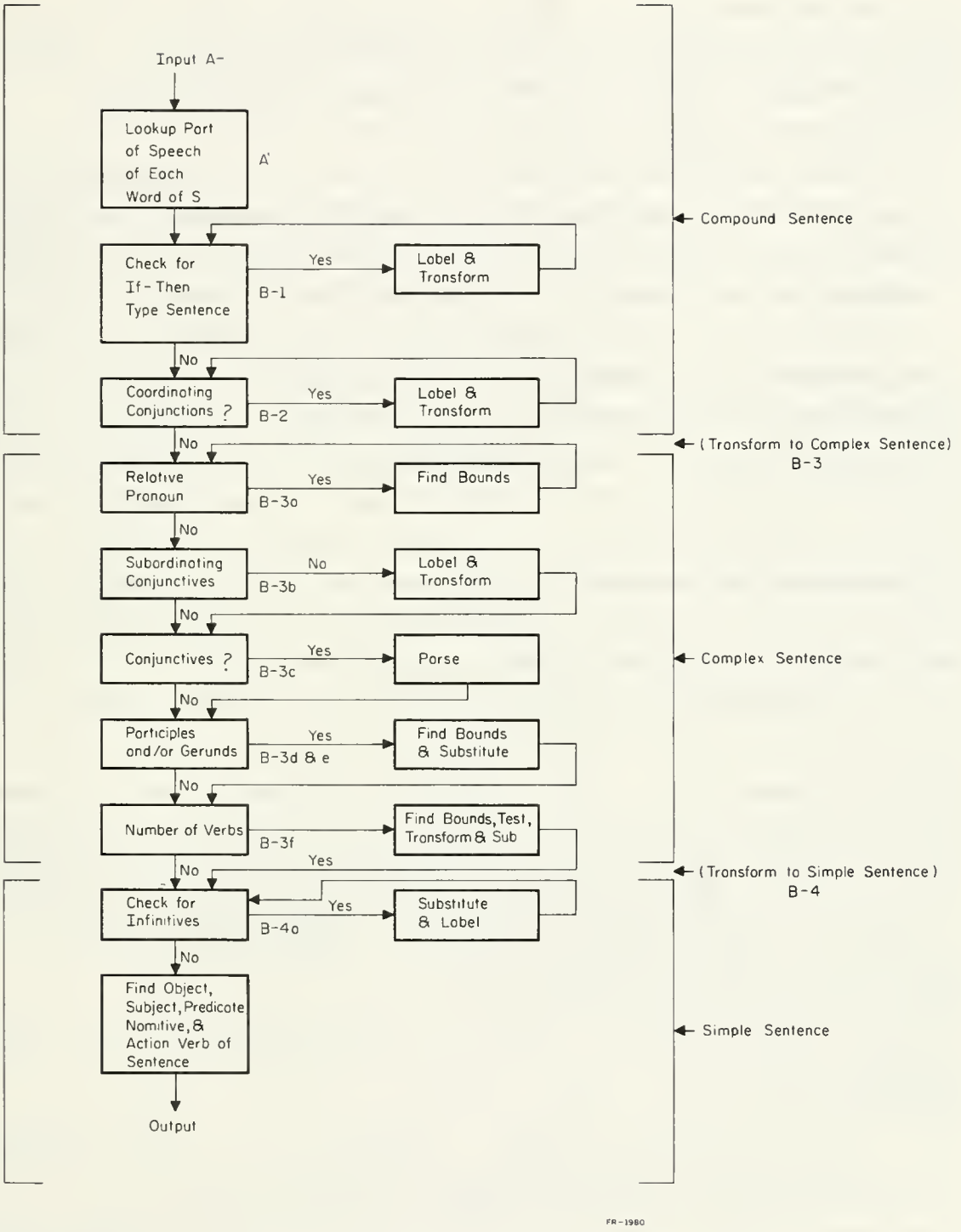


Fig. 15.2. Flow chart for parser.

if..., then..., or ..., if..., or if..., ..., or ..., then...
If one of these conditions occur, then we send the program into a transformation routine. If it does not occur, then we continue to B-2.

B-2. Check for coordinating conjunction

In this part of the program, we look at each coordinating conjunction and see if it is the connective or two independent sentences. We do this for each coordinating conjunction until it either works or we are finished in checking all the coordinating conjunctions. The procedure is the following:

We look at the first coordinating conjunction and see if there is a comma (phrase-marker) immediately following the coordinating conjunction. We do this for all the conjunctions (coordinating). If this does not produce a transformation into two independent sentences, we check the environment of each coordinating conjunction. This is done by first looking at the left side and comparing its structure to the right side, since it is a coordinating conjunction and left-hand structure must equal the right-hand structure. If the structure is a sentence, then we can make the necessary transformation. If not, we store the structure for later reference. (In the checking of the environment of the coordinating conjunction, we first go to the left-most boundary which is either another conjunction or a series of blanks which means we are at the beginning of the input string. The same is done on the right-hand side if it is going to be transformed.) We continue this checking until we have exhausted the number of coordinating conjunctions.

If we cannot transform S into two complex sentence S_1' and S_2' then we transform S into the complex sentence S', and continue to B-3.

List of coordinating conjunctions are:

(and, for, but, a nor, yet, both... ... and, not only... at also, either.... or, neither...nor)

B-3a. Check for Relative Pronouns

The pronouns in the Relative Pronoun list are relative pronouns if the phrase-marker "?" does not appear in the input string.

1. We check the input string to see if first there exists a relative pronoun. If one exists, then we find its environment and we make the necessary transformations.

The environment of a relative pronoun is almost always a simple sentence. The environment is found by looking to the right of the pronoun,

and by immediate constituent analysis we find the simple sentence in the environment. After this is found, we label the structure R and substitute it into the original sentence. We now look at what classification R can be. It can either be used as a noun or as a modifier. This can be handled by the simple test to see if a NP is needed to make the complex sentence have syntactical sense. It can either be subject, object, or object of a preposition if it is a NP. If there already is an object of a sentence, or a subject of a sentence, or no hanging prepositions, then the relative pronoun and its environment will be used as a modifier. It will modify either the subject or the object (or predicate nominative) depending upon which it is nearest to. When we decide what the relative pronoun's use is, we give R its equivalent value, either NP or Aj.

After this test we proceed to B-3b.

List of relative pronouns:

(who, which, what, that, whoever, whatever, whichever, whom, whomever)

B-3b. Check for Subordinating Conjunctions

If there is a subordinating conjunction in the complex sentence, we make a transformation as found in our transformation rules:

$$\ldots, SC\ldots \rightarrow SC(S_1', S_2')$$

$$SC\ldots, \ldots SC(S_2', S_1')$$

This is done by finding the subordinating conjunction and looking to the right of the conjunction until we get a series of blanks, if the SC is not the first word of the sentence. We label the SC's environment as S_2 and the words to the left of the SC as S_1 and perform the necessary transformation. If the subordinating conjunction is at the beginning of the sentence, we must find its environment by the complex-sentence test (minus subroutine C.2.) and immediate Constitutent analysis. When we find its environment, we perform the necessary transformation. After this test we proceed to B-3c.

List of subordinating conjunctions: (although, though, that, because, since, so that, in order that, as, unless, before, where, when).

Check for Conjunctives

B-3c. This is a simple test. If there is a conjunctive, it is set off in commas, both on its right and left sides, and, since it is adverb, it modifies the main verb of the sentence. After this test, we proceed to B-3,d,e.

List of conjunctives:

(however, furthermore, therefore, moreover, nevertheless)

B-3d & B-3e. Check for Participles and Gerunds

The word V_{ing} can either be a participle (a verb form used as an adjective) or it can be used as a gerund (a verb form used as a noun) or it can be the main verb of the sentence (if it is used with the verb to be).

We first test to see if it is used with the verb to be, since this is the most simplest form of the V_{ing} verb. It must follow immediately after the verb to be or the word not, to satisfy the previous relation.

If it does not satisfy the condition of being a verb along with to be, then we will assume it is a gerund. This will imply that it must be the object of the sentence, subject of the sentence, or object of a dangling preposition. If it fits into one of these categories we are finished. We will label the V_{ing} G and substitute it into the complex sentence. If it is neither a verb or a gerund we will call it a participle, and since, it is an adjective, it will modify a noun, which will follow either immediately after or two words after the participle.

A participle can also be a verb in the form Vad. A participle and a gerund have an environment. These are the words (or word) immediately following a participle or gerund. The word (word and modifiers) is the object, predicate nominative, or predicate adjective of the Ving of Ved form. (This happens because a participle or a gerund still have properties of being a verb.) Therefore, the participle and gerund will have a syntactical analysis also. After this test we go to B-3f.

B-3f. Check for more then 1 verb

If there are more than 1 verb in the sentence, (V+V constructions mean 1 verb) implies that S' still a complex sentence. Verbs taken care of in coordinating conjunctions, subordinating conjunctions, and relative pronoun phrases are not counted. This usually occurs when a relative pronoun is left out. If this occurs, we should place a relative pronoun in front of the first verb (second, third,...if necessary) to see if it makes sense and thus eliminates the problem of extra verbs. After this test we go to B-4 then to B-4a.

B-ra. Check for Infinitives

Now all we have left is a simple sentence and various substitutions.

However, there may be an infinitive in the simple sentence. The form of an infinitive is the following: to verb. An infinitive may be used as a noun, adjective, or adverb.

The procedure is to first check to see if it is a noun, the same way we checked if a relative pronoun is a noun. If it is not a noun,and it follows directly after a noun or an adjective, it then modifies the noun or adjective it follows,and then we can substitute AD for the infinitive. If it is not a noun, and it follows directly after a verb, then it modifies the verb, and it is labeled an adverb. After this test we go to B-4b.

B-4b. Simple Sentence S".

We are now left with a simple sentence and various substitutions. This can easily be handled by immediate constituent analysis, and simple transformations.

Table 1. Transformation Rules

1. CC + , - CC(S_1',S_2')
2. S + CC + S - CC)S_1',S_2')
3. NP + CC + NP = NP
4. VP + VP = VP
5. V + CC + V = V
6. N + CC + N = N
7. N + CC + NP = NP
8. NP + CC + N = NP
9. S + SC + S = SC(S_1',S_2')
10. SC + S + S = SC(S_2',S_1')
11. SC + NP + S = SC(S,SCP)
12. ...

Table 2. Grammar Rules

1. N + N = NP
2. Aj + N = NP
3. D + N = NP
4. D + NP = NP
5. Aj + NP = NP
6. Aux + V = VP
7. P + NP = PP
8. VP + PP = VP

9. NP + PP = NP
10. Ad + V = VP
11. NP + VP = S
12. S + NP = S
13. ...

J. Schultz

15.1.4. Logical Structure and Semantic Modeling

The general theory of "looking-up"a preprogrammed answer for the input question is highly unsophisticated and both time and space consuming. A design that has a built-in logic structure using atomic concepts is planned.

The cellular structure is based on the logical concepts of the ACTION WORDS, and one is able to picture the scene by looking at the structure. For the ease of machine implementation and manipulation, mathematical functions are formed.

There are approximately five such cells for the ACTION WORDS: yield, pass and lane usage, turns and signals, speed restrictions, stopping and parking. To illustrate, the ACTION WORD "turn" is chosen. The cell is able to handle reasonable questions such as: "A driver is going north, and wants to turn right. What lane should he be in before entering the intersection? After the completion of turning, what lane should he be in?"

The atomic concepts for this particular ACTION WORD "turn" are: initial direction, final direction, initial lane position, final lane position. They are not totally dependent on the input question and thus avoiding the fallacy of "matching the right answer to the right question."

A cartesian coordinate system is set up with the x axis corresponding to the initial cardinal direction of the vehicle. The final direction can be calculated by the function. Furthermore, if given the initial and final directions, the type of turn can also be calculated. Hence the logic structure is well defined.

On calculating the proper initial and final lane usage, the following logic is used. If a driver is on a four-lane road and wants to turn left he should be in the left lane before turning. After completing the turn, he should also be in the left lane in the appropriate direction.

This logical strategy is attempted for the five ACTION WORDS for traffic law.

T. Woo

15.1.5. Theorem-Proving Techniques

In the past several months, we have continued our study of automatic theorem-proving algorithms and we are now trying to apply these techniques to the R2 system.

At this point we are working on a representation of the drivers manual in the First-order-predicate calculus so that we can apply the theorem-proving algorithm that we have studied.

Kenneth Biss

15.1.6. Semantic Structures

Two areas were investigated during this reporting period. The first area is concerned with determining a model of a computer memory structure for storing English text. This structure would be used in conjunction with a computer program in order to "answer questions" concerning the stored text. The structure itself is built up from two types of elements--called nodes and links. A word from the text is associated with either a node or a link via a pointer from a dictionary. The words, as such, are not stored as part of the structure.

Nodes are normally referenced by pointers from nouns, pronouns, adjectives, and adverbs, while links are usually associated with verbs and conjunctions. A link is used to tie together two nodes. The model allows a node-link node relationship to be viewed as a single node thus allowing for a recursively defined structure.

The second topic investigated is that of formatted-file techniques. Examples of the words in this area include examining the connection between fields in line items, determining techniques which show explicitly how to evaluate the worth of secondary indexes, and finding methods to simulate formatted file usage.

E. A. Mark

15.2. Associative Processor Architecture

In the past year, a study was conducted which aimed at finding a quasi-optimal architecture for an information-retrieval processor. This study, will be presented as a Coordinated Science Laboratory report soon. Some late developments and a conclusion will be given here.

One of the problems of a large processor designed for information retrieval is the partitioning of a set of I/O channels and processor cells into blocks. (Each cell contains storage facilities for a processor word, a comparator, and some control circuitry in the stored program interative processor.) Each block then can contain enough I/O channels and cells to execute a program or answer a query. Separate blocks can then operate independently to "space share" the processor. A solution to this problem is the SW-structure (switching structure) described below.

The SW-structure is a connection network that is derived from a tree. In Fig. 15.3, all of the tree above level 2 is reproduced and attached to the original tree at level 2 to obtain Fig. 15.4. In that figure, a tree structure can be restored by disconnecting links between levels three and two. In fact, two tree structures can be obtained. For example, by cutting links between nodes (3,2) and (2,2) and again between nodes (3,4) and (2,4). One tree has nodes (3,4) (2,2), (1,1) and (1,2), and the other tree has nodes (3,2), (2,4), (1,3) and (1,4). In Fig. 15.5, all of the tree above level 1 has been reproduced again. Here, four tree structures can be obtained. In fact, for each partitioning of the nodes (1,1), (1,2), (1,3) and (1,4), a collection of disjoint trees can be found such that each tree has the nodes of a block of the partition. In essence, then, an SW-structure is a connection network that partitions a set of nodes, here at level 1, into blocks, and provides a tree structure for communication between all members of each block.

This process of reproducing can be done at each level of a tree having ℓ levels. It can be generalized to a tree whose fanout is f, just as easily as to the binary tree of Fig. 15.3. Further, more than one reproduction, say s-1 reproductions, can be made at each level. The resulting structure is a (uniform) SW-structure of ℓ levels, with fanout f, and spread s. This structure will have $f^{\ell-1}$ base nodes (at the bottom of the structure) and $s^{\ell-1}$ vertex nodes (at the top of the structure). By means of

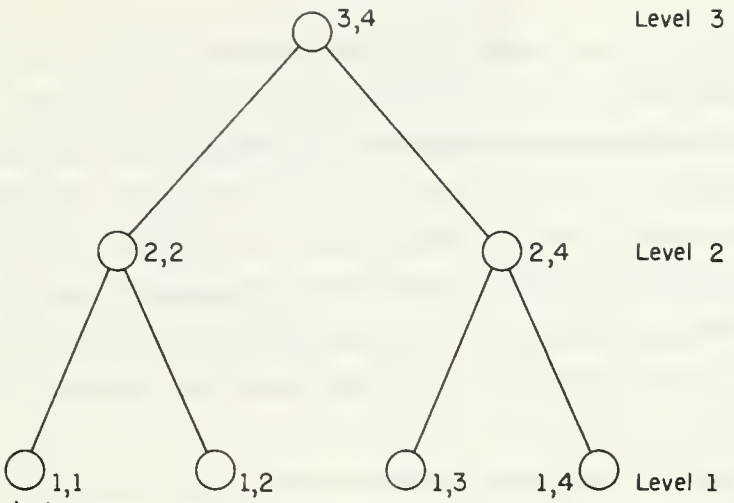


Fig. 15.3. A tree.

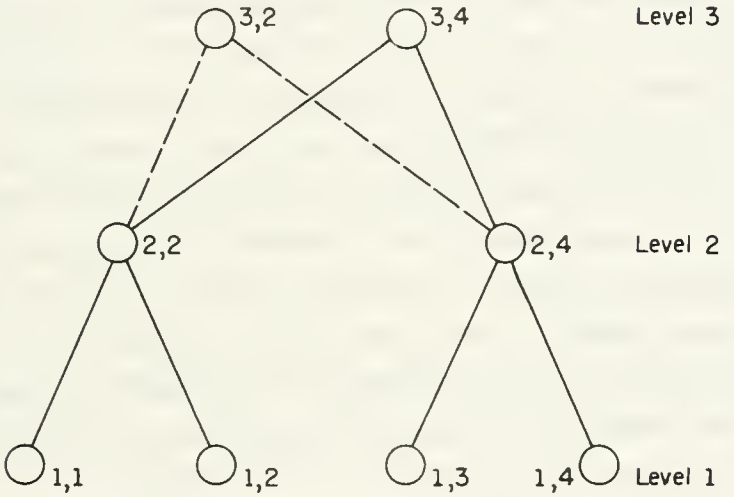


Fig. 15.4. One reproduction.

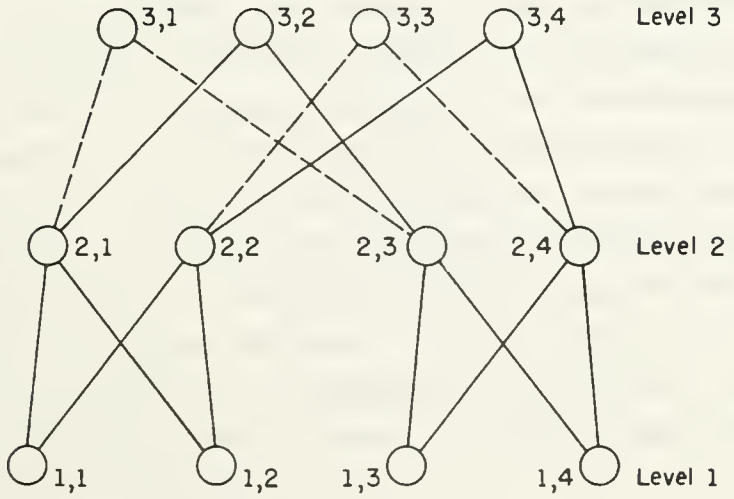


Fig. 15.5. Two reproductions.

such an SW-structure, the set of $f^{\ell-1}$ base nodes can be partitioned into $s^{\ell-1}$ or fewer blocks.

For the intended application, each base node will be connected to either an I/O channel or to a segment of an essentially linear array of cells. The SW-structure will concatenate these segments to produce a larger linear array of cells suitable for processing. This structure has low propagation-delay times, it can function in the presence of faulty cells, and it can be efficiently realized in a regular way in three-dimensional space for some cases.

The aims of the study conducted over the past year have been attained. A processor designed for information retrieval must be large to be efficient, but large processors have been generally slow and unreliable, and they require a kind of "space sharing" to be efficient. In this study, a tree structure (not the SW-structure above) was found in which theoretically 10^6 cells can be used with a clock period of one-half microsecond. The structure is amenable to a fault-detection technique and can operate in the presence of faulty cells. This structure can itself be divided into independently acting subtrees, or several trees can be connected together by an SW-structure, to vary the effective size of the processor. Each independent subtree is a stored-program iterative subprocessor. All these features can be added to a conventional associative memory cell at a "cost" increase of approximately 20%.

An instruction set was obtained for this processor which is faster for information-retrieval problems, yet less "costly", than those proposed for other processors. A linked descriptor structure was given which is a suitable format for storing information in this or other associative memories. A technique using pushdown stacks was found useful in generating instructions in this stored-program iterative processor. Finally, a description of a format for a decision function was given; this function produces a sequence of instructions in response to a binary state vector which is broadcast into the processor. The time required to select the instruction sequence can be made essentially independent of the complexity of the mapping from the binary state vector to the instruction sequence.

J. Lipovski

15.3. Investigation of Perceptive Cognition

The general philosophy of our project on perceptive cognition, tentatively outlined in the two previous progress reports, has been further refined and will be reviewed in this section. We feel that our current understanding of the problem is mature for the algorithmic implementation on a computer of a simple pilot experiment, which we plan to initiate before the next reporting date. As a programmatic rule, we have adopted the view to reduce the complexity of the problem to a minimum level, compatible with the clear retention of cognitive elements (as opposed to purely classificatory). In this fashion, it is conceivable that more complex situations can be handled by simply extending the system under development.

The input to the cognitive system S is represented by landscape-like scenes (although this choice bears no character of necessity). The expression "landscape-like" is meant to indicate the general context from which inputs are drawn and the presence of some important deviation from, say, a photograph of the landscape. These deviations are aimed at simplifying the input by way of a deterministic preprocessing, that is, not involving any activity of cognitive nature. The first step of visual preprocessing is the reduction of the scene to a "cartoon" of itself. The scene is subdivided according to a uniform grid, each of the resulting little domains is classified according to, say, its color, then slight contrasts are removed and strong contrasts are sharpened. This leads to the use of separation lines between areas of relatively uniform physical characteristics. These areas, denoted hereafter as domains, are labeled D_1, D_2, ... D_s. The second step is the classification of each D_j according to a small set of elementary attributes a_1, a_2, ... a_b, (such as size, color, etc.). The third step is the verification of few very simple geometric relations existing among the D_j's, such as inclusion, adjacency, vertical ordering, etc. For each examined relation R we can then produce a directed graph where $D_j R D_i$ is represented by a directed link from node D_j to node D_i. If a relation is an equivalence its graph becomes a lattice. Each node of the mentioned graph is further characterized by an attribute vector.

Example: If the relation R is "inclusion", then a typical example is given in Fig. 15.6. The top node is the whole scene, the bottom node is the "empty element" and all intermediate nodes are the D_j's. It is clear that the

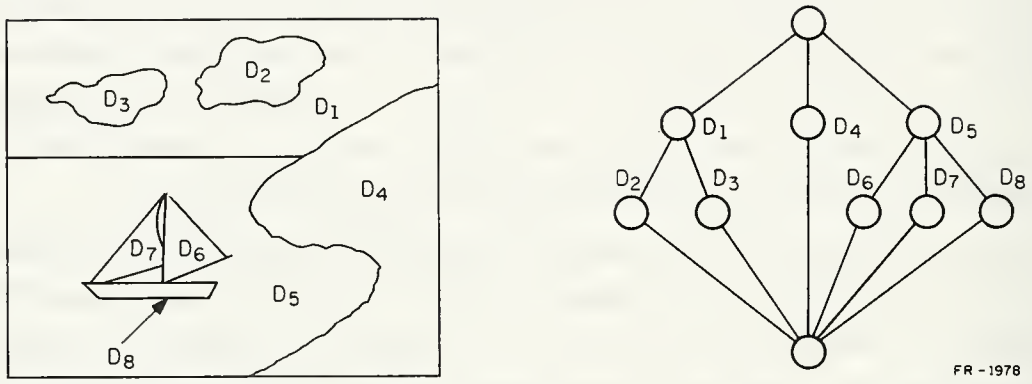


Fig. 15.6. A simple scene and its descriptive graph for the relation of inclusion.

mapping of an input scene into its descriptive graphs is a purely mechanical activity not involving any "guess" or interpretation of the scene according to a stored model. This leads us to operational definition of "cognition" of visual inputs.

By cognition we mean the process of mapping a visual input into a description or interpretation. The process of input interpretation is very grossly analogous to the operation of decoding in statistical communication theory, with one substantial difference. Both situations are characterized by the existence of "equivalence classes" of the inputs which are to be mapped into the same "statement". But while in communication theory the equivalence classes are determined by the statistical properties of noise, in cognition they are determined by the practical necessity to be able to encompass a very large number of different visual inputs with a single description. The outstanding common feature of decoding and cognition is their being decision activities, in that they are designed to produce the most likely or reasonable interpretation of their input. This analogy is far reaching and provides the philosophy of our algorithm for cognition.

An essential consequence of the previous discussion is that the decision process requires comparing the given input against a stored "image" of the structure of the events to be recognized: This is usually referred to as the world-model or eidology.* Cognition is achieved when there is sufficient evidence in support of an interpretation of a given input against any other interpretation. This chosen interpretation, termed the description, is expressed and communicated as the labeling, according to a conventionally established "lexical" code of the following items:

(1) Context and perceivable objects of the scene

(2) Inferred (nonperceivable) properties of the scene.

It should be noted that the distinction between 1) and 2) is not one of nature, but simply one of degree of difficulty. For this reason, the first

---

*It is interesting to note that a fundamental prerequisite of visual cognition is that the recognizable be an extremely small subset of the visually representable. A similar situation appears to hold between the cardinality of the descriptions and of the described events.

version of the proposed system S will concern itself primarily with item (1) and very modestly with item (2).

Although conceivable in principle, a cognitive procedure based on the global access to all the details of the scene in order to produce an instantaneous scene-description mapping is clearly computationally unfeasible, unless the universe is rudimentarily simple, e.g., a graphic alphabet, the arabic numerals. In fact the unfeasibility derives from the fact that the input must be compared for agreement with all possible descriptions. Whence the idea of an algorithm necessarily sequential in nature, which discards very early the unreasonable interpretations and retains only the most likely candidates for success.* It is apparent from the previous discussion that the sequential algorithm implies proceeding from general to specific. The most obvious general-specific pair is certainly whole-part or object-detail. Since, also very frequently, whole and part satisfy the inclusion relation, the inclusion lattice is seen as a very powerful aid in carrying out the cognitive procedure. In rough terms, the underlying philosophy can be summarized as follows: ·"Some evidence is used to formulate a hypothesis on some object; the hypothesis is tested by verifying the presence of some expected details." With these premises, the algorithm for cognition can now be outlined.

The process is characterized by the movement of two pointers, the object pointer and the model pointer. The former scans the set of the graphs describing the various geometric relations and dwells at any one time, on homologous nodes in these graphs. The latter moves within the stored image of the events, denoted as archive, and indicates with its position the heretofore accumulated cognition. In order to clarify the subsequent discussion, we now briefly sketch a possible format for the archive.

The archive is composed of records, labeled after objects $0_1, 0_2, \ldots, 0_m$. Each record 0_j is characterized by a set of attributes $\{a_i\}$

*The stated principle is also the basis of sequential decoding of convolutional codes. The algorithm shares in fact several fundamental properties with sequential decoding, as we shall see.

and by a set of contextual specifications of the form $R_h O_k$ (R_h is a geometric relation). The former set is meant to specify the physical properties of O_j and the latter the most reasonable characteristics of its environment.

Cognition results from the interaction of the two pointers. A "forward look" of the model pointer represents for formulation of a hypothesis O_j for the node under examination. This hypothesis provides the object pointer with a set of tests, in relational form, to be verified. Hence, movements of the object pointer are executions of tests or extraction of attributes. If the tests provide sufficient evidence in support of the hypothesis O_j, then the hypothesis is confirmed, and becomes an element of the "cognized set". The model pointer is now moved to O_j. The basic component of the algorithm is functionally sketched in Fig. 15.7. The object pointer is moved to the "next" node (the strategy for the sequencing of the nodes will be described later), and the node attributes are collected (step 1). This evidence is used in performing the comparison of the extracted attributes against the attributes of the set of objects which are "expected" at this point (step 2). The key-point is that this set, being conditioned by the heretofore recognized context, is likely to be very small (for example, the set of major details of a given object). This suggests the conjecture that comparison 2 could be performed as linear classification, since the number of outcomes is small, and the various outcomes should have sufficiently different characteristic. This would revive the viability of procedured "a la perceptron" at an incremental level (as opposed to global).

In the assumption of successful comparison, i.e., the attributes are classified as those of some object O_j, the hypothesis "O_j has been observed" is formulated (step 3). This hypothesis is now available for verification. The record pertaining to O_j is accessed and the test to be performed are obtained (step 4). These tests are then executed, typically by checking whether the node under examination is related in specified ways to other nodes of specified attributes (step 5). If this verification is successful, O_j is confirmed and retained as acquired evidence (step 6).

Clearly, the crucial steps of the described procedure are those which involve deciding whether comparisons (steps 2 and 5) are or are not

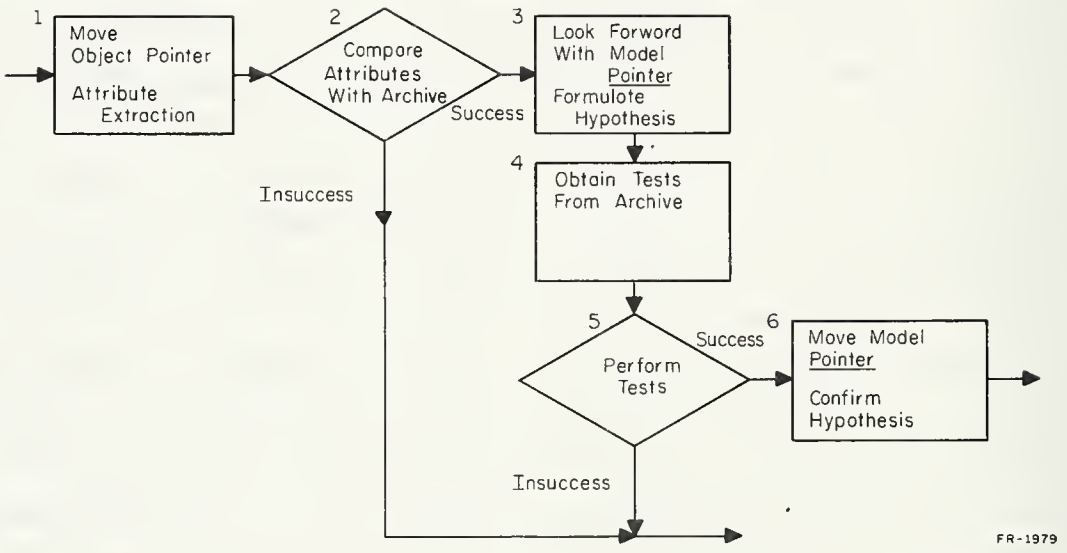


Fig. 15.7. Basic operational component of the cognitive algorithm.

successful. This entails the definition of a scoring function φ and of an acceptance criterion θ. Specifically, given a hypothesis O_j we evaluate $\varphi(O_j)$; if $\varphi(O_j) > \theta$, then O_j is formulated. It appears therefore that $\varphi(O_j)$ must be indicative of the "reasonableness" of O_j within the context so far discriminated. Assume that in the cognitive process we have successfully discriminated objects $O_{h_1}, O_{h_2}, \ldots, O_{h_m}$. Then in order to express the contextual dependence, it must be

$$\varphi(O_j) = \varphi(O_j | R_{k_1} O_{h_1}, \ldots, R_{k_m} O_{h_m}).$$

In other words, φ appears to be analogous to a likelihood function in a statistical decision system. An even more convincing analogy is obtained by considering φ as the mutual information between the set $O_{h_1}, \ldots, O_{h_m}$ and the object O_j related to O_{h_i} by relation R_{h_i}. A probability measure could be assigned on the resulting sample space, as representative of the constraints existing within sets of objects. This, however, appears to lead to an intractable complexity. To avoid this difficulty, such a probability measure could be defined only on the list of relations contained in each record (see the preceding description of the archive). This probability, or association measure, should be incorporated, a priori, in the system at least for the most general levels of the cognitive process. This would allow the capability of accomplishing at least some rudimentary contextual recognition. It is conceivable that the structure required for a more detailed cognition could be incrementally added to the system through an adaptive procedure modeled as a teacher-student interaction. The investigation of this aspect, however, is temporarily deferred.

We now consider in some detail the strategy for node sequencing. In harmony with the criterion to proceed from general to specific, the visual preprocessive step organizes objects in hierarchical levels according to their apparent (i.e., determinable, for example, on the basis of the inclusion relation) whole-part relationship: objects so organized are therefore in a predecessor-successor vertical hierarchy. Among all the successors of a given predecessor a new horizontal hierarchy is obtained by ranking the objects according to some function of their attributes (e.g., size). Then the object pointer will proceed descending in the given hierarchies, with the vertical hierarchy predominating over the horizontal one.

Finally, we consider a point which has been intentionally omitted so far. It is possible that the evaluation of φ leads to a hypothesis O_j which is false. This falsity will become manifest in evaluating φ for the expected details of O_j and receiving negative evidence (insuccess exit lines in the flow-diagram of Fig. 15.5). It will be possible, however, to correct for the erroneous interpretation by moving backward the object pointer (backtracking) until φ is again above the threshold for acceptance. This entails replacing O_j with the "next best" hypothesis. The procedure then continues unaltered.

F. P. Preparata
S. Ray

15.4. Perception and Cognitive Structures

15.4.1. Perceptual Foundation of Linguistical Meaning

An experiment was designed to elucidate the role of perception in the acquisition of cognitive structure. Linguistical meaning was chosen as the cognitive structure for investigation. No attempt was made to characterize the notion of "meaning". Rather, the experiment was so designed as to allow the delineation of pertinent characteristics of this phenomenon, which could be used in an appropriate characterization.

The experiment consisted of three sessions. At the beginning of the first session, a subject was read the following instructions: "I am going to read a list of nouns. I want you to describe the thing named by each noun. Please make your descriptions as complete as possible." The list of nouns contained words from a dozen categories, for example, fruit, flowers, animals, etc. Three situations were anticipated: (1) The subject had perceptually encountered the thing named by the noun; (2) The subject had not perceptually encountered the thing named by the noun, due to historical circumstance; (3) The subject had not perceptually encountered the thing named by the noun, due to perceptual inability. For example, a congenitally blind person could not have perceptually encountered a rainbow; he may, or may not, have perceptually encountered a radish.

The second session occurred 4 days after the first. At the beginning of the session, the following instructions were given: "One

object at a time will be placed before you. Please investigate it in any way you choose and as much as you desire. After exploring the object to your satisfaction, indicate whether you have encountered the object previously." For those nouns whose denotations could not be presented, other nouns were substituted.

The third session occurred 4 days after the second. At the beginning of the session, the same instructions were read as at the beginning of the first session. The list of nouns read consisted of the list of nouns presented in the first session together with the substituted nouns in the second session.

The experiment was conducted with a pilot group of five subjects, three congenitally blind and two partially sighted. All the subjects were college or graduate students. All sessions were tape recorded.

The data is presently being analyzed. An analysis totally depending upon frequency counts of parts of speech does not seem appropriate, since such an analysis does not measure the appropriateness of a subject's usage of a lexical item. An analysis typical of the Piaget School seems more useful. In the Piaget approach, a transcript of the experimental session is presented and the important phenomenon contained therein are analyzed.

A preliminary analysis indicates that perceptual encounter is a necessary condition for a subject to gain total "meaning" for a noun. Some subjects are capable of giving an extensive and appropriate description of the thing named by the noun, but are not able to recognize the thing named by the noun when it is presented to him. Subjects sometimes state that they really do not know to what their descriptions refer.

The experimental technique and design is being refined. A second group of blind subjects will be tested, as well as a control group of sighted subjects.

15.4.2. A Perceptual Foundation of Dreams in the Blind

It is expected that some cognitive structures are derived from and dependent upon perceptual processes. To test this conjecture, a perceptual phenomenon typical of the blind was chosen for this experiment.

A blind person after self-training or after special education can learn to perceive objects by means of hearing. The reflection of ambient

noise from an object is a sufficient cue for the perception of that object. As a blind person approaches an object he perceives the object as "looming" toward him. No details about the objects other than their size can be discerned.

Since this sense is a primary source of perceptual information for a blind person, it would be expected that he would inculcate such perceptions into his cognitive structures. The blind subjects were asked to describe how objects appear to them by means of this sense. When the subjects were asked to describe the perceptual content of their dreams, they invariably included the "looming" sensation in their descriptions. It is apparent that this perceptual phenomenon can be readily represented in cognitive structures of the blind, in particular, in their dream structure.

15.4.3. A Theory of Result Systems

An algebraic system is a nonempty set of objects together with one or more operations on that set. An operation is a process which combines a given number of objects into a unique result. The result of combination is assumed to belong to the nonempty set of the algebraic system. There are many examples of structures in cognitive psychology, biology, and anthropology which do not quite satisfy the requirements of an algebraic system. In particular, the result of combining a given number of objects in the given nonempty set may not belong to the nonempty set. Instead, the result of combination may only satisfy some properties typical of the members of the given nonempty set.

A new mathematical system can be defined which generalizes the notion of an algebraic system. A result system consists of a nonempty set of objects, a nonempty set of predicates, and a nonempty set of processes. Each process combines a given number of the objects of the nonempty set of objects in to a unique result. This result possesses a subset of the set of predicates. Result systems are being developed and their utility investigated.

15.4.4. Paradoxes of Logic and Set Theory

Some attempts at modeling the paradoxes of logic and set theory have been made. Models of these paradoxes may prove useful in the development of result systems. Both involve predicates, and both may use self-referential constructions involving predicates.

P. Duran

15.5. Semantic Compiler

There are in existence today many compilers of various types for the current babel of programming languages. They are united by the fact that all work strictly with the syntax level of the language, finding in the input text an allowable string of some type and then transforming it into code as a unit independent of its surroundings. This approach works well if the program contains at most syntactic errors which the compiler can recognize as such. If, however, the program contains logical error, if it is a false algorithm, so to speak, then the results are less satisfactory. Therefore, work has started on the application of ideas gained from the study of natural language, and others, to the processing of computer programs. It is felt that the result could be a higher-level debugging aid which would detect a class of logic and design errors outside the capabilities of present compilers, thus forming the basis for an interactive man-machine program-composition system.

From the work already done on this, the following conclusions have been made:

(1) that a programming language represents an implicit definition of a pseudomachine (that is: instead of an IBM 360 or a CDC 1604, a FORTRAN computer or an L6 computer) of which machine language is only the most specific case;

(2) that a program is a sequence of relationships between the states of this machine;

(3) that one can represent these relationships (via cylinder-type data structures) in a way which will allow them to be algorithmically manipulated;

(4) that in an interactive system of this type it should not be necessary for the user to have extensive knowledge beyond a familiarity with the programming language he is working with, although some type of (presumably pseudo-English) metalanguage for statements about program segments will be necessary.

A program is now working which reads programs written in ILLAR (the current assembly language on the 1604) and produces data structures representing all of their internal relations of control flow and memory reference. These data structures correlate both with the source program

records and with the binary version of the program and can be used either for interpretive execution (for debugging), or for assembling a normal binary version of the program. Currently under way is the editing feature which accepts editing directives referring to the source program, appropriately modifying the data structure, which automatically updates both source and binary programs. When the editing feature is finished, the interpreter will be written. The data structure contains enough information to allow the design of an interpreter which will detect many errors too subtle for an assembler, and to provide information to the user indicating the probable causes of the difficulty in some detail. Among detectable errors would be many cases of program looping, transfer of control to now programs areas, storage in nondata areas, and misuse of data type, i.e., mixture of fixed-point and floating-point operations on the same data, use of floating numbers as addresses, etc.

S. Taylor
P. Weston
V. Wu

15.6. Data Structures and Interactive Programming

A technical report (R-393) describing the CYL6 ring-structure routine system has been prepared and released. This CYL6 system was described in the previous Semi-Annual Progress Report.

Employing CYL6 in the management of the required data bases, a group of four interactive programs, collectively referred to as EPIXYS, has been designed and is now in operation. This group consists of LEJARDIN, a graphics program which allows light-pen controlled drawing of patterns on the display screen, plotting of arbitrary numerical data, and the grouping of picture elements into hierarchical structures. Each subgroup in a picture has its own name and is related to its supergroup by a coordinate transformation allowing rotation, translations, and scaling of the subgroup independently of other groups. In addition, each subgroup, which in the data structure is represented by a list of its contained picture elements and subgroups, may also contain one or more iteration elements. These elements have the function of specifying the repetitive plotting of all following items in the list. The iterative plotting is under the control

of six parameters; the iteration count, the horizontal and vertical displacements of each subsequent replot, i.e., a displacement vector, a scale factor to be applied cumulatively to each replot, a rotation also applied cumulatively, and two more, a scale factor and rotation which are applied only to the displacement vector and not to the plotted material. The iteration elements simplify such ordinary drafting applications as constructing graph grids and tic marks, and the drawing of polygons, but also open the door to some interesting abstract designs as well.

Control of LEJARDIN is accomplished through a tracking cross and light buttons on the display and a command language available through typewriter or paper- or magnetic-tape input. The latter features allow the design of graphic materials off line and use of the interactive facility for making final alterations without the need for repeated runs to reach a desired result. The magnetic-tape control feature makes possible control from another program and this facility is used in the second program in EPIXYS, THEODOR.

Because the CSL display is equipped with both hardware and software to facilitate the composition of display images into movie sequences, and to transfer these to film, we have provided a program to provide an integrated movie control facility for LEJARDIN. THEODOR has two input languages. In the first, which is not available on-line, screen action is described in terms of the coordinates, scales, and iteration parameters of specified LEJARDIN picture structures. These specifications are organized into units referred to as scenes and given individual names. A group of scenes may be used under the control of the second on-line language governing the construction of control tapes which feed frame-by-frame directions to LEJARDIN, allowing it to generate a movie tape which can later be run through a standard system routine to produce a movie film, or may be previewed through the display screen. The THEODOR on-line directives include the running of a given scene for a given time or frame count, rewinding the film to a given point or by some frame count, fading in, or fading out. The user is informed after each on-line directive about the present footage and running time of his final film.

The third member of the EPIXYS group is SAM, a program which allows on- or off-line specification of functions and their on-line evaluation or plotting. The organization of the data structure in SAM is highly parallel to that in LEJARDIN. That is, all functions in the system at any time are individually named and are represented by a list of subfunctions chosen from among the other current named functions or from the set of available elementary functions. The latter currently includes sin, cos, log, exp, absolute value, integer power, square root, noninteger power, and a random generator. The system also includes a list of variables and constants, supplied by the user, which are used as arguments to the elementary functions, and which may be designated as argument variables to the user-defined functions.

User-defined functions are of only two basic types, sum or product. For example, a fully expanded polynomial would be represented by a sum function containing a number of subfunction terms of the product type. Each term in a function list is marked with a "+" or "-" sign, indicating whether the value of that term should enter the value of the function through the direct or inverse operation. That is, the value of an element bearing a "-" sign is subtracted from the value of a sum function while the value of a product function is divided by such a term. Complete generality is obtained by including one more type of function element marking, which facilitates function composition. For this type, the element or subfunction is evaluated as usual, but instead of combining its value with the partial function value, it is instead stored as the value of a specified variable, which in turn may be used as the argument of a subfunction appearing later in the element list. For example, the function

$$\cos(\log(x))$$

could be specified by a two element list, the first of which computes log(x) and puts this value back in x, the second computing cos(x).

As construction of function definitions proceeds on-line, the display is used to show the user the current contents of the function he is defining, along with the contents of all involved subfunctions, and, with this feedback, on-line use has been made reasonably efficient. Display of this information also provides for rapid recording of function definitions on the Polaroid camera unit.

The value of any function for any particular values of its arguments is available on-line, as is a plot of any function over a specified range of any one of its arguments, for given values of the remaining arguments. The number of points to be plotted is also specifiable. The ordinate values of any such plot may be output on various media for subsequent use by other programs or for use as permanent records.

It is the use of curves generated by SAM which occasioned its addition to EPIXYS. Such data are useable either directly in LEJARDIN as plotted data forming picture elements or indirectly, in constructing scene definitions for THEODOR, which in turn control LEJARDIN picture-parameter values. SAM does, however, have other possibilities because it is faster and more flexible overall than FORTRAN programming as a means for quickly getting data on the characteristics of functions, e.g., the zeros of transcendental functions, or the behavior of multivariable functions in certain regions of their domains. With the addition of an iteration-element type and a test element the capabilities of SAM can be extended to functions containing differential or integral expressions as well as recursive functions.

The final member of EPIXYS is MUSIC, a program which allows the output of complex waveshapes to external amplifying or recording equipment. Presently the waveshapes are designed only on-line under light-pen control, and have certain stringent form constraints, but curves generated by SAM or any other source could also be made available. Output control is through the typewriter keyboard presently, which for this purpose is interpreted to have a certain frequency and waveshape associated with each key, rather than an alphanumeric character.

Because the EPIXYS programs each fill the core of the 1604, but are also capable of transmitting considerable quantities of data, they are provided with magnetic tape as a communication medium and with a typewriter-controlled facility to preserve the detailed state of any of them, including its data structure, then to restart any other selected program in the system from a previously preserved state.

S. Taylor
P. Weston
V. Wu

15.7. Mathematical Modeling of Interactions in Neurons and Nets

It is extremely important to develop a physiologically realistic model of a neuron, realistic at least in the sense that it is not restricted to linear summation of inputs. Neurophysiologists have known for some time that one cannot speak of neurons as a single class of almost-identical elements. Neurons obviously differ greatly in their structures and functions; their responses to subthreshold stimuli; facilitation; fatigue properties--the list is almost endless. For this reason any model that is claimed to be a model of THE NEURON must be viewed with suspicion; upon examination it will almost invariably be found to be badly flawed.

In attempting to construct models of neural elements, one is faced with a compromise between, on the one hand, extreme realism for a very specific class of cells [7], and on the other hand, rather broad modeling of general properties which will usually not withstand comparison with any specific cell class. If it is decided to construct a realistic model for a particular cell class, it will generally be found to be impossible to treat the model mathematically and the only recourse will be to simulation or to the construction of physical realizations of the model (usually electronic [8]). This approach, the modeling of specific cell types, is analytic in nature, in that it is most often used as an adjunct to neurophysiological investigation, as a tool to generate predictions and to test hypotheses about the preparation under study. The alternative approach, that of broad modeling of general properties, is usually more amenable to mathematical investigation; however, it is generally far harder to generate powerful predictions and significant hypotheses with this approach. Being more synthetic in nature, the modeller is more often than not attempting to construct artifacts that duplicate some of the properties of living neural systems, and is more concerned with overall 'input-output' behaviors than with the details of internal functioning.

The model under consideration is synthetic rather than analytic in nature; it deals with a broad class of phenomena in neural systems, emphasis is given to interactions in dendritic fields of a neuron and to overall information transmission by neurons. However, it also appears that the model can be applied successfully to some specific classes, the simplest of which

are lateral inhibition nets and other nets that compute simple invariants such as convexity and linearity. Further, the theoretical framework being developed has great promise for applications to pattern recognition in an engineering context.

The essentials of the model under development are relatively simple and are summarized in the following paragraphs.

The basic element of the theory is the unit, later to be identified with the neuron. Each unit possesses an input field (IF), corresponding to a neuron's dendritic field. The signals converging on the IF are pure frequencies, which for convenience are normalized to the range 0 to 1. No distinction is made between inhibitory and excitatory signals, as these characteristics appear only in the IF. Each afferent to an IF terminates at a particular point in the IF (a synapse), where the signal in terms of frequency f_i is transformed into a signal in terms of another variable X_i, $X_i = X_i(f_i)$. The variable X_i may be thought of as a post-synaptic potential. The functional form of X_i is left unspecified, as it can be specified later without introducing any severe mathematical complications. Restrictions placed for simplicity on X_i are that X_i be a single-valued increasing function of f_i, and that $X_i \geq 0$.

$X_i(t)$ depends only on $f_i(t)$, no temporal summation being introduced into the elementary theory. If desired, simple temporal summation can be introduced by writing $X_i(t)$ as a convolution of f_i with some exponential weight function. But in keeping with the viewpoint current among many contemporary neurophysiologists, time is neglected in the preliminary version of the theory, and 'transfer functions' at one particular instant of time are examined in detail.

There are two classes of X_i in the IF, excitatory and inhibitory. Excitatory X_i, written E_i, are additive as is usually assumed in neural modeling. Inhibitory signals, N_i, are multiplicative. This is in keeping with the viewpoint that inhibition is a relatively nonselective shunt across the dendritic membrane, (Maturana, H., personal communication). The net effect of the E_i and the N_i on a single segment of the IF is of course completely dependent on the geometry of the IF--where the excitatory and inhibitory terminals lie in relation to one another and to the unit itself. It is impossible to separate the function of the IF from its structure.

All signals X_i produced in the IF are collected at one point, the root (or soma). Figure 15.8 gives some illustrative examples of simple (and probably unrealistic) structures and the associated equations.

The basic assumption regarding the operation of inhibition is as follows: Inhibition by N_i at a point P on a segment of IF has the effect of multiplying all signals above P, away from the root, by a function of N_i. Relatively little appears to be known about the precise functional form of this function, but the following choice appears to be reasonable both on mathematical grounds and in the light of physiological experience. Let all signals above P, away from the root, be multiplied by $(1-nN_i)$, where n is a scaling factor such that max $nN_i=1$. (Recall that N_i is maximum for maximum input frequency f_i). An alternative choice that has not been explored is $(1+nN_i)^{-1}$, which reduces to the above for small nN_i. Without evidence to the contrary, the former is the expression of choice on grounds of simplicity. However, it is well known that in the actual neural membrane, the relation between the amount of transmitter released and hyperpolarization is linear, at least for small amounts of transmitter, and small rates of change of transmitter release. Linearity also holds between the average instantaneous frequency and the amount of transmitter release, for small f_i and f_i.*

As stated above, no temporal effects are considered and the signals in the IF are assumed to propagate instantaneously through the IF. This is a questionable assumption, since we know that inhibition can be extremely long-lasting in its effects compared with excitation (e.g., in cortical pyramidal cells). However, as a first approximation, it is a reasonable mathematical simplification, and it appears that time lags can be introduced later with relatively little trouble. (Intuitively the introduction of time lags would be satisfying, as Reiss has shown that a system operating with time delays as the information carrier can carry out quite complex computations).

---

*One relatively simple model of transmitter release and polarization has been constructed during the present research, specifically to check the assumptions of linearity.

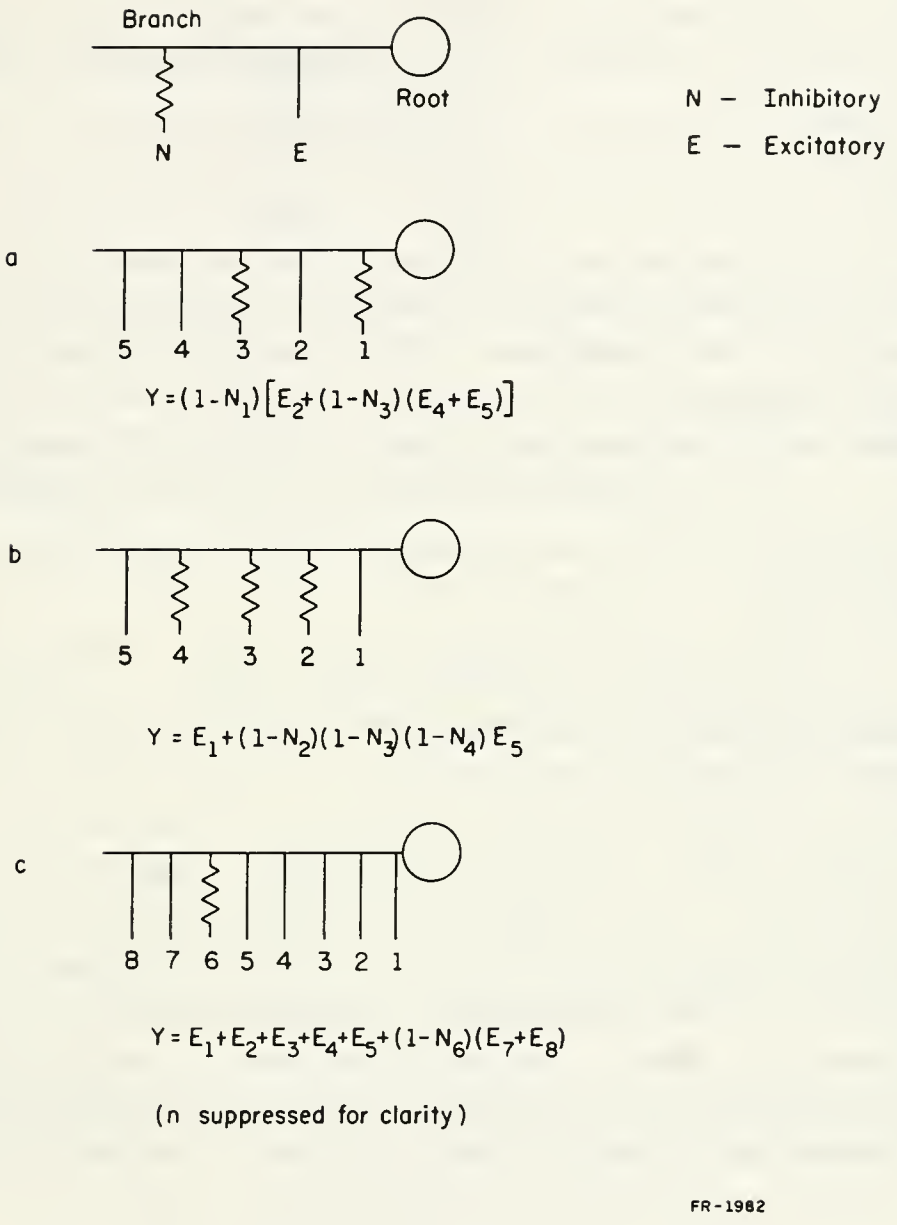


Fig. 15.8. Neural structures.

One of the major stumbling blocks in neural modeling has been the presence of the highly nonlinear threshold of neurons. However, in many cases, and especially for relatively long input-pulse trains, the nonlinear threshold can be replaced by a linear or logarithmic characteristic. This may be done without any reference to the processes prior to the soma. In the present work attention focusses on interactive processes prior to the soma, allowing the threshold to be replaced by an unspecified function. If, at a later time, the function requires specification, a linear or logarithmic function will be used. The use of the full nonlinear threshold will undoubtedly require the introduction of simulation techniques.

If the equations for the representative structures in Fig. 15.8 are expanded, one sees that the general form of the equation for one branch is as follows:

$$y = \sum_i E_i - \sum_i N_i \sum_{j>i} E_j + \sum_{ij>i} N_i N_j \sum_{k>j} E_k - \sum_i \sum_{k>j} N_i N_j N_k \sum_{e>h} E_e + \ldots \pm \prod_i N_i \sum_{j>i} E_j$$

The important fact to note here is that inhibition occurs in partial products, excitation in partial sums.

In earlier reports, reference was made to interaction space. This abstract space is constructed as follows. Suppose the unit has n inputs $(x_1, x_2, \ldots, x_n)$. Construct all products of the form $x_i x_j$, $x_i x_j x_k, \ldots, x_1 x_2 x_3 \ldots x_n$ where $i>j>k\ldots$. Now form the space with 2^n-1 axes, the axes being the original n inputs and all possible products like the above; this space is the interaction space. A vector in this space, with 2^n-1 components c_i specifies a possible state of the units IF. The input to the root, and the output of the unit in the present model, is

$$\begin{aligned} z = {} & x_1+x_2+\ldots+x_n+x_1x_2+x_1x_3+\ldots+x_1x_n \\ & +x_2x_3+\ldots+x_2x_n+\ldots+x_{n-1}x_n \\ & +x_1x_2x_3+\ldots+x_1x_2x_3\ldots x_n. \end{aligned} \tag{2}$$

A direct treatment of this space is difficult mathematically, although some simple results can be obtained by trivial applications of vector analysis.

Because of mathematical difficulties associated with this interaction space, another approach was sought. Examining Eq.(1), one notes that the inhibition terms alone, neglecting excitation, occur in products just like the products used in the interaction space. Consider the column

vector $(E_1,E_2,\ldots,E_k)$ representing all inputs to an IF that are excitatory. Apply the following linear transform:

$$\underset{\sim}{E}' = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & 1 & 1 & \ldots & 1 \\ 0 & 0 & 1 & \ldots & 1 \\ \vdots & & & & \vdots \\ 0 & 0 & & \ldots & 1 \end{pmatrix} \underset{\sim}{E} \; . \tag{3}$$

The column vector $\underset{\sim}{E}'$ contains all the partial products that enter into Eq.(1). The mathematical difficulties are still present in the various partial products of the N_i's, but the problem can be simplified by the introduction of 'polycorrelations.' The mth-order polycorrelation among a set of n inhibitory signals N_i is defined as follows:

$$R_m(\tau_1,\ldots,\tau_m) = \sum_i \sum_{\tau_1>i} \ldots \sum_{\tau_m>\tau_{m-1}} N_i N_{i+\tau_1} N_{i+\tau_2} \ldots N_{i+\tau_m} \; , \tag{4}$$

that definition being by analogy with ordinary correlation, and by a slight generalization of the terms entering into Eq.(1). The 0-th-order polycorrelation is defined as

$$R_0 = \sum_i N_i \tag{5}$$

Note that $R_1(t)$ is just the ordinary autocorrelation. $R_k(t_1,\ldots,t_k)$ generates all (k+1)-th-order interactions that appear between the N_i in interaction space.

It is possible to establish a connection here with expansions of a Boolean-valued variable. If one restricts all inputs to the values of 0 or 1, and all operations are carried out mod 2, then the set of all interactions between n inputs--the interaction space elements--is just the finite Walsh expansion less unity [9]. This expansion is one of the standard methods for expanding a Boolean-valued function of binary variables, and is closely related to the Rademacher function of combinational analysis, and the Lazarsfeld-Bahadur expansion. With respect to the inner product

$$\langle z_i, z_j \rangle = 2^{-n} \sum_{\{x\}} z_i z_j = \delta_{ij} \tag{6}$$

the finite Walsh function is a CON set. This suggests that in the interaction-space representation, and in the analog expression for a unit, Eq.(1), unity should be added. In fact, unity is the coefficient of the total E sum in Eq.(1), although it does not appear explicitly. The problem now is that of finding a useful and valid inner product in inhibitory interaction space, hence defining a norm. This will make the inhibitory interaction space a normed linear space, thus opening the way for a fully-rigorous mathematical treatment and avoiding the earlier difficulties experienced with interaction space. In this context it should also be noted that Eq.(1) is also multilinear in the N_i's.

Earlier in the year, some elementary tools of tensor calculus and differential geometry were applied to interaction space, although there was still doubt about the mathematical validity of the methods. Now that a CON set can be constructed in interaction space, these methods and tools can be freely applied. This is especially important in that tensor calculus allows the concise and powerful treatment of complex structures in the space. The total behavior of a unit may be viewed geometrically as follows: form the space of 2^n components, consisting of the interactions and original input, Eq.(2), together with the output, here considered as a scalar. Then the total behavior of the unit, under all inputs, is a hypersurface in this space. A specific behavior is a trajectory in this neural hypersurface. The various problems stated earlier in the year can now be formulated in precise fashion; of especial interest being the variational formulation of the trajectory equations. Just as in differential geometry, geodesics can be computed by minimizing a quadratic functional; can the neural trajectories in the neural hypersurface be defined by minimizing an appropriate functional? This problem is of great interest, given the "Principle of the Lazy Neuron" developed earlier. This principle may be stated as follows:

> In sufficiently cephalized neural systems, the behavior of a central neuron, or a sufficiently closely connected net, is subject to the following constraint. The total information exchange during a computation is a minimum consistent with the computation.

A computation, in the sense used here, is the response of a unit to a specific pattern of input $(x_1,x_2,\ldots,x_n)$, and is therefore a point in the neural hypersurface. A quantity J, the information exchange during a computation, is associated with every point of the neural hypersurface. The above principle requires that J be a minimum everywhere on the hypersurface, consistent with the computation. This last phrase means that the unit responds in some selective fashion to the input vector; for example, if the unit is an edge detector, it must yield a maximal response to all edges in its IF, irrespective of their orientation or scale, or position, and a lower response to all other structures presented to its IF. It is appropriate to note here that the polycorrelations introduced earlier automatically remove translations from the output, they are translation invariant. Concerning the nature of J, it is certainly not the classic Shannon measure, and its determination is one of the objects of the current research. One can express the principle of the lazy neuron in variational terms, using the computation performed as a constraint; then, knowing J, one could in principle determine the structure of the neural hypersurface. However, J is unknown, so the problem has to be inverted, and the question asked, given a particular hypersurface structure, and a particular computation, what functional J could give rise to these? This is a well-defined problem, although the resulting expression for J will be arbitrary to a certain extent, and further data will be needed to determine J in any given situation.

The importance of determining J cannot be overstated. Given J, Noether's theorems [10] can be applied. These theorems provide a very powerful and elegant way of computing invariants of the system under given sets of transformations. Suppose that a set of r-parameter transformations (or even r-function transformations) is applied to the neural hypersurface. The resulting transformation of the hypersurface can be regarded as a change in the computation of the unit, or a change in the function being computed. If J is required to be invariant under the transformation--in other words, the principle of the lazy neuron holds before and after transformation--then Noether's theorem provides a method for computing a set of invariants for the system. Invariants are important quantities

for any system, and furthermore are nearly always observables*. It should be noted here that a change in the neural hypersurface can maybe be viewed as a plastic change in the unit, and it therefore seems possible that such elusive phenomena as 'memory' may be amenable to treatment in the theory presently being developed. Any transformation applied to the neural hypersurface can, as is well known, be regarded from two points of view, the alias and alibi. In one viewpoint, the surface is regarded as being changed, and this is interpreted as a plastic change in the unit. In the other viewpoint, it is the coordinates that are regarded as being changed, and this appears to be equivalent to supplying the unit with a new set of inputs--in other words, placing the same unit in a new invironment.

It may be thought that the introduction of the polycorrelation functions is a complicating factor in an already complex situation. However, by the use of multidimensional Fourier transforms, the polycorrelations may be reduced to very simple expressions. There is some work available on the use of higher-order autocorrelations in pattern-recognition work, especially the recognition of cursive characters [11], and there are indications that such functions are a powerful tool. The appearance of similar functions, the polycorrelations, appears to indicate the way to a unified theory of pattern recognition and neural behavior, areas that have been felt by many workers to be closely related. Bernhard [12] has proposed a theory of essentially continuous perceptrons as a model of neural nets, and, in his theory, activity is represented by propagating waves of first-order correlations. He based his theory on quantum mechanical analogies, while the present theory, which leads to polycorrelations, is more firmly based on physiology.

It was noted earlier that the function computed by a unit is entirely dependent on the geometric structure of the IF of the unit.

---

*The question of important parameters and observables for a neuron is an important and unsolved question. Associated with any neuron there are certainly important quantities other than its input-output transfer function for all inputs, but the nature of these quantities is not yet known.

This leads to the occurrence of residual terms, if the polycorrelation expressions are substituted into Eq.(1) for a particular structure. The treatment of these residual terms appears to offer some difficulty, but this can be circumvented if the following intuitively satisfying postulate is made:

> To a high degree of approximation, real IF's (dendritic trees) are so structured that the residual terms are either entirely absent, or are completely dominated by the polycorrelations.

In other words, real neurons are polycorrelation computers. From the synthetic viewpoint of course, the IF's will be designed with no residual terms, as pure polycorrelation computers.

A few significant properties of the structures under development are as follows:

(1) A unit would appear to be resistant to damage or perturbation of inhibitory inputs, as these occur in all possible products, and a change in one of the N_i's will not drastically affect the sum of products. Excitation would appear to be more sensitive to perturbations.

(2) Correlation has been used with great success in optomotor studies [13]. These studies were in systems terms, the only physiologically identifiable elements being the input and output, yet they met with great success. The present approach offers a way of extending these and similar studies down to the neural level.

(3) In a gross behavioral context Greene [14] has suggested that organisms carry out computations akin to factor analysis, or automatic diagonalization of the input covariance matrices. This leads to the concept of an organism expressing input stimuli in terms of an internally generated orthogonal basis or set of references, thus enabling comparisons to be made with sets of internally generated or stored goals, and internally stored past stimuli. It appears that the units being studied theoretically in the present work may be the units with which to realize such a scheme.

P. Weston
M. Wilkins

15.8. References

1. Siklossy, L., Natural Language Learning by Computer, Department of Computer Science, Carnegie-Mellon University, Pittsburgh (1968).

2. Woods, W. A., "Procedural Semantics for a Question-Answering Machine" in Proceedings FJCC, pp. 457-471 (1968).

3. Simmons, K. F., J. F. Burger, R. M. Schwarcz, "A Computational Model of Verbal Understanding" in Proceedings FJCC, pp. 441-456 (1968).

4. Schwarcz, R. M., J. F. Burger, R. F. Simmons, Systems Development Corporation Report SP. 3272 (1968).

5. Robinson, J. A., "A Machine-Oriented Logic Based on the Resolution Principle," JACM, 12, 4, pp. 23-41 (1965).

6. Robinson, J. A., "Heuristic and Complete Processes in the Mechanization of Theorem Proving" in Systems and Computer Science, University of Toronto, pp. 116-124 (1967).

7. MacGregor, R. J., A Digital Computer Model of Spike Elicitation by PSP's in Single Nerve Cells, RAND report RM-4877-ARPA, Santa Monica, California.

8. Lewis, E. R., "Using Electronic Circuits to Model Simple Neuro-Electric Interactions" Proc. IEEE 56, 931-949 (1968).

9. Ito, T., "A Note on a General Expansion of Functions of Binary Variables," Inform. & Control 12, 206-211 (1968).

10. Gel'Fand, I.M. and S. V. Fomin, Calculus of Variations, Prentice-Hall, New Jersey (1963).

11. McLaughlin, J. A. and J. Raviv, "Nth-Order Autocorrelations in Pattern Recognition," Inform. & Control 12, 121-142 (1968).

12. Bernhard, R., "Correlation Waves in Brain-Like Structures," Bull. Math. Biophys. 27, 435-447 (1965).

13. Reichardt, W., "Nervous Processing of Sensory Information," in Theoretical and Mathematical Biology, T. H. Waterman and H. J. Morowitz (eds.), Blaisdell & Co., New York, pp. 344-370 (1965).

14. Green, P. H., "An Approach to Computers that Perceive, Learn and Reason" in Proc. of Wester Joint Computer Conference, pp. 181-186 (1959).

16. SEISMIC VEHICLE DETECTION[†]

B. D. Kirkwood
H. W. Knoebel
J. Gooch
V. C. Metze

16.1 Spectral Analysis

In a continued effort to evaluate the feasibility of the identification of vehicles by their frequency-ratio signatures, an attempt is being made to determine which components of the vehicle's seismic signal are associated with the engine rotation rate. As a first attempt, one wheel of a vehicle was jacked up and the distributor breaker-point signal was recorded along with the seismic signal produced while the vehicle remained in place with the engine running and one rear wheel rotating. The breaker-point signal was used as a clock for digital sampling of the seismic signal for spectral analysis. Although the wheel rotation rate was evident in the spectra, the resolution expected was not achieved. It appears that the breaker-point signal is a poor clock due to the action of the vacuum- and centrifugal-advance mechanisms as well as wear and contact bounce. The next step was to provide a better clock signal and a telemetry system to permit taking data with the vehicle in motion.

A magnetic pickup was installed on the flywheel housing to produce pulse as each tooth of the flywheel gear passes. By dividing the pulse frequency with integrated-circuit dividers, a range of sampling rates are available which are stationary with respect to crankshaft rotation.

A pair of Handie-Talkie trancievers has been modified to serve as a telemetry link for the flywheel clock signal to permit recording engine clock data with the vehicle in motion. Additional work is needed on an external telemetry antenna for the truck. A few records have been made but the telemetry signal becomes very noisy whenever the present antenna is screened by the truck body.

16.2 Weighing of Moving Vehicles

A need exists for a system for catagorization of moving vehicles by

[†]This work supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199.

weight. In traffic counts for evaluation of highway wear, it is desirable to discriminate between automobiles and trucks on a weight basis. For arms control inspection, the United States Arms Control and Disarmament Agency has a need to detect and catagorize by weight both wheeled and tracked vehicles.

A few preliminary tests have been made of the use of geophone signals for weighing passing vehicles. These tests show some of the promise and some of the limitations of this technique.

If an ideal accelerometer is buried under a road surface and its output integrated twice, the road surface deflection due to a passing vehicle could be determined and in turn the weight of the vehicle ascertained. Since the output of a geophone above its resonant frequency is proportional to velocity, it should, for high frequencies, give road deflection with a single integration. Below resonance, three integrations would be required. For these initial tests, the geophone signal was integrated once before recording on magnetic tape as shown in the block diagram of Fig. 16.1.

Typical waveforms of automobile and truck signals before and after integration are shown in Fig. 16.2. At low speeds, each axle is resolved. The difference in the loading of the front and rear axles of the truck is evident. With the geophone underdamped (approximately 30% of critical damping) the axle signals start to merge as the time separation, at approximately 3 m/s (7 mi/hr), becomes less than the natural period of the geophone (1s) and appear unresolved at 9 m/s (20 mi/hr), Fig. 16.4. With the geophone near critical damping, the axles still appear to be resolved at 6.7 m/s (15 mi/hr) as seen in Fig. 16.3. More data are clearly needed at higher speeds. The amplitude behavior as a function of speed needs more investigation also. Most of the tests have been made in a parking area surfaced with crushed rock where the space available limits the maximum speed to about 9 m/s (20 mi/hr). The signal has been observed to be similar with the geophone sitting on the edge of a paved street, but none of this data has been recorded.

Figure 16.5 shows the result of a few low speed tests with a 1967 Chevrolet station wagon on the parking lot when the ground was frozen to a depth of 0.6m. After some slight thawing, another test was run including an eight-ton truck with dual rear wheels. The results, shown in Fig. 16.6, give a plot of peak integrated signal amplitude as a function of distance from the

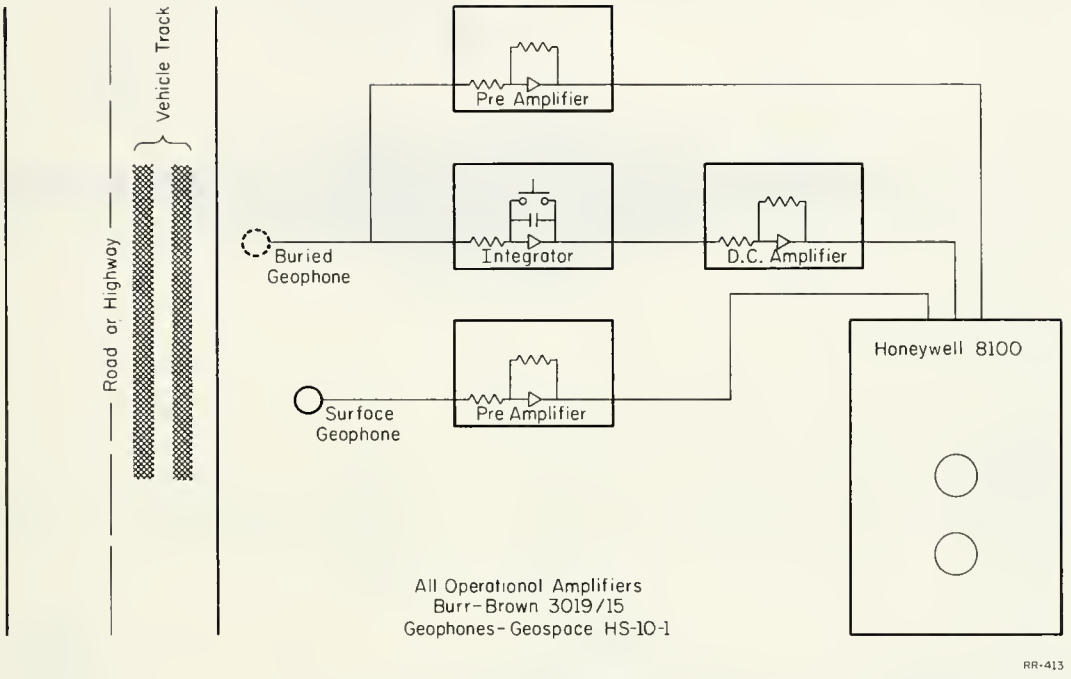


Fig. 16.1. Block diagram of instrumentation for recording integrated seismic signals.

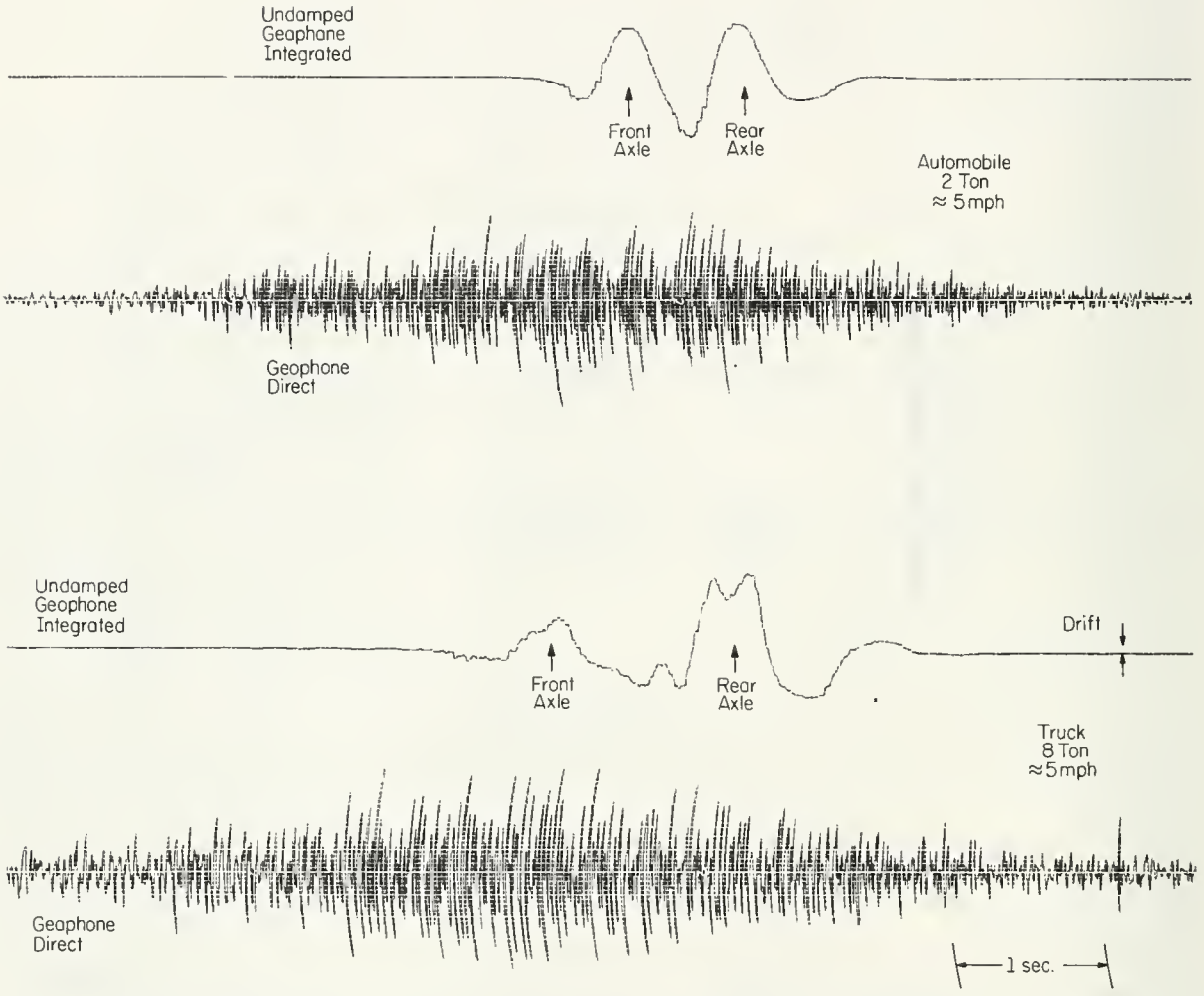


Fig. 16.2. Typical waveforms of direct and integrated geophone signals from an automobile and a truck.

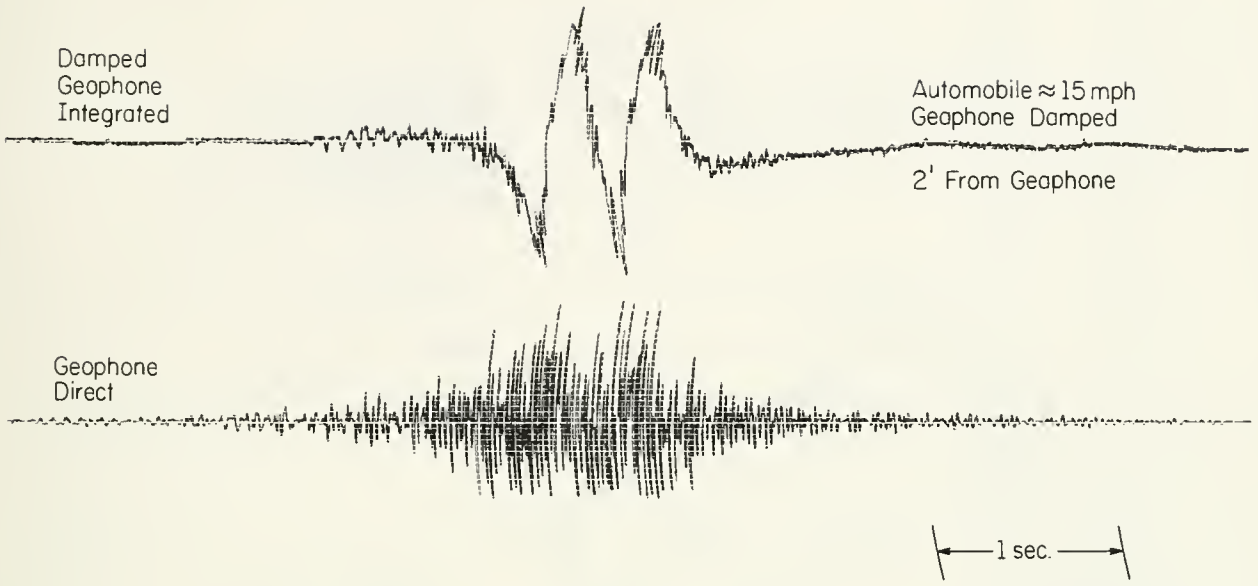


Fig. 16.3. Direct and integrated geophone waveforms with geophone near critical damping.

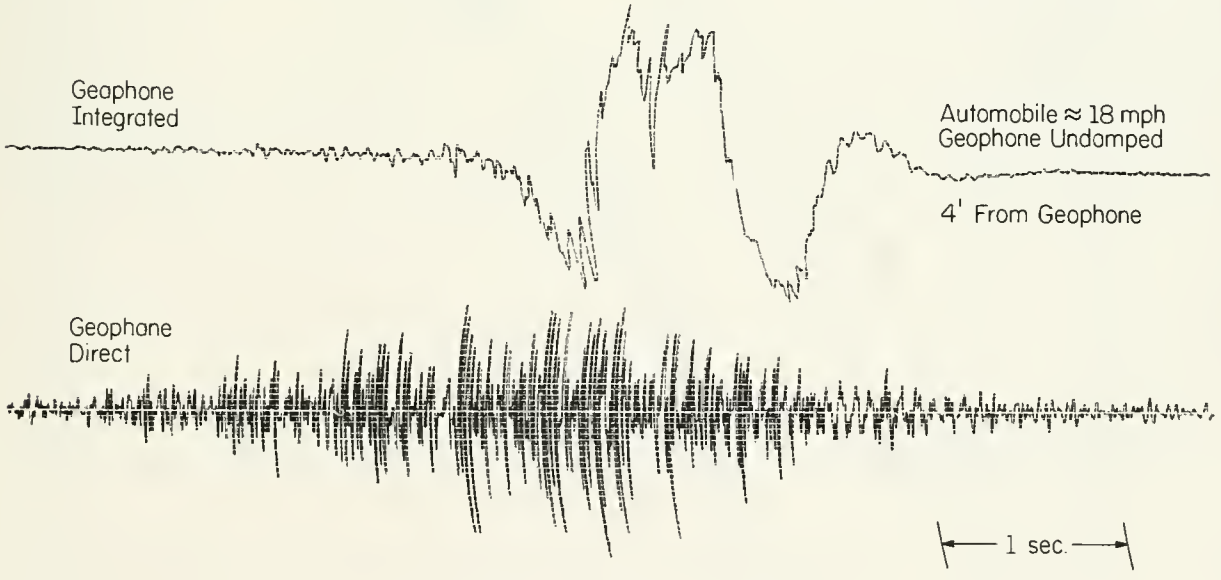


Fig. 16.4. Direct and integrated geophone waveforms with geophone underdamped.

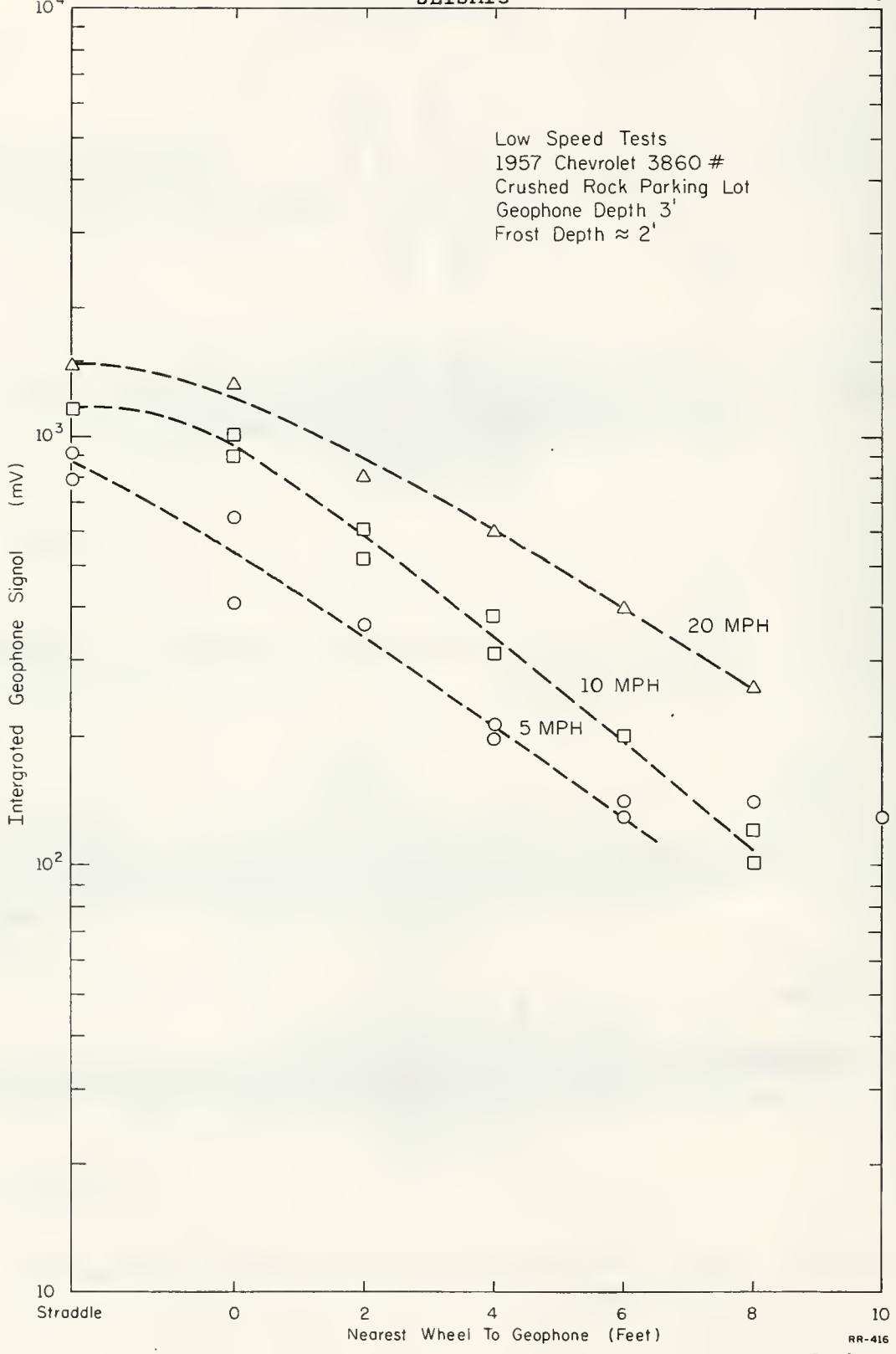


Fig. 16.5. Integrated geophone-signal amplitude as a function of the distance from the vehicle to the geophone for a 1957 Chevrolet.

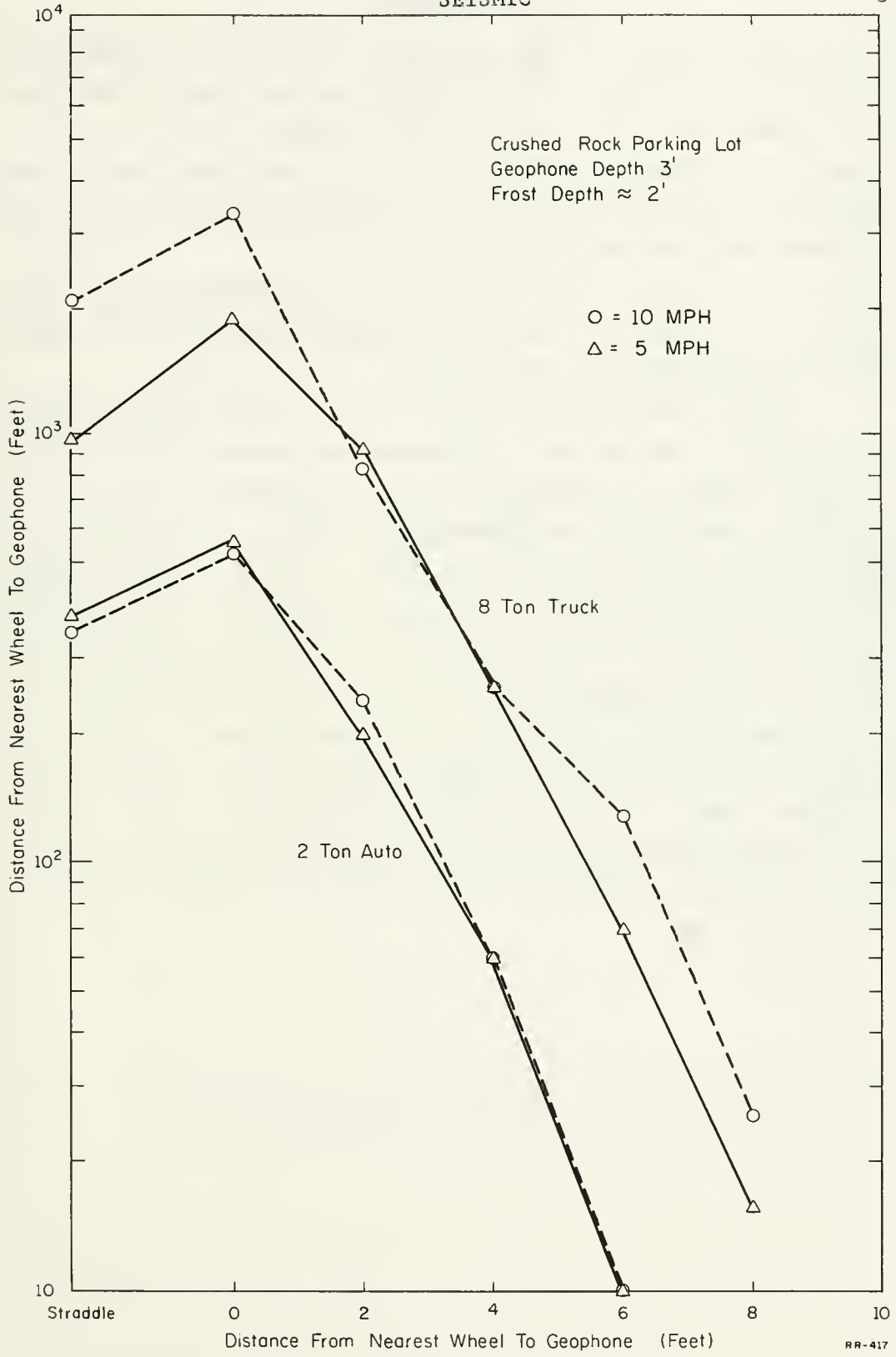


Fig. 16.6. Integrated geophone-signal amplitude as a function of the distance from the vehicle to the geophone for a 2-tone automobile and an 8-ton truck.

geophone to the track of the nearest wheel. As before, the signal falls off rapidly with distance (note the logarithmic vertical scale). The behavior in the region where the vehicle straddles the geophone differs in the two tests. This may be due to the frozen layer acting as a structural beam in the tests of Fig. 16.5 while in Fig. 16.6 the heavy truck has broken up the remaining frozen earth which now acts as a plastic mass.

Additional recordings have been made of National Guard vehicles including a jeep, 3/4-ton truck, 2½-ton truck, armored personnel carrier, and a tank. These data have not been analyzed.

With the limited data analyzed so far it appears that, for categorization by weight on an unpaved road, sensors would need to be buried more than 0.6m deep. (Tests with the geophone on the surface and buried 0.3m gave very erratic data). The system would need speed information either self-generated (geophones along the direction of travel) or provided by other sensors. As an aid to determining the number and spacing of geophones, the graph in Fig. 16.7 was generated by approximating the data of Fig. 16.6, folding it about the geophone location as though the vehicle in various runs had passed first on one side of the geophone, straddled it, and then passed on the other side. The sum of the outputs of two of these geophones spaced 3m apart gives the curves shown. Under these very restricted conditions, one could discriminate between a two-ton and an eight-ton vehicle over at least a 4.3m roadway. For more precise weighing, more closely-spaced sensors may be required.

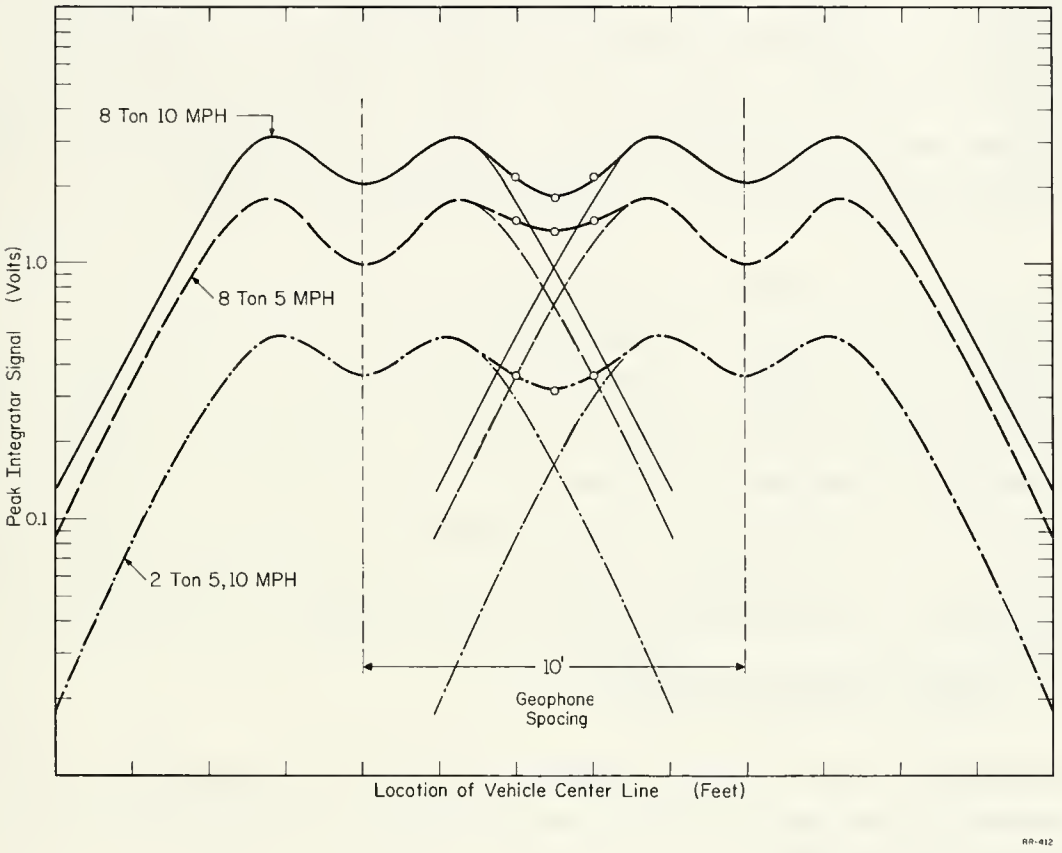


Fig. 16.7. Extrapolation of single-geophone data to estimate the weight resolution of two geophones spaced 3 m (10 ft) apart.

17. ENGINEERING AIDS FOR THE HANDICAPPED†

H. W. Knoebel
J. G. Burr
G. Stupp

17.1 Introduction

A new program for the development of systems, with the emphasis on information-processing systems, which will enable improved communication and better control of their environment by handicapped individuals is being started. The control of highway vehicles by quadriplegics (individuals with severe impediment of both arms and both legs) and techniques enabling blind to read ordinary newsprint are two problems currently under study.

The University of Illinois has the oldest and largest rehabilitation program of any university in the nation. Most classrooms and teaching laboratories are accessible to students in wheelchairs. Special buses provide wheelchair transportation. Handicapped students are able to compete academically and move about the campus with the same independence as able-bodied students. This rehabilitation program provides a wealth of subjects with which to work in the development of new systems and also makes evident the severity of the transportation problem faced by handicapped graduates taking employment in a world designed with little consideration for their needs.

17.2 Vehicle for the Disabled

Work has been initiated on the development of a highway vehicle operable by a severely disabled person. The vehicle is based on the concept of a senior design project [1] done in the spring of 1968 in the General Engineering Department.

A trip has been made to the General Motors Research Laboratory in Warren, Michigan to investigate their work in fully-automatic automobile control systems. Observations were made of the modifications made in standard automotive hardware. Their experience in the design of control systems was found to be of great value in shaping the course of our work.

Through the efforts of the General Engineering Department, General Motors is donating a Chevrolet step van for use as a basis for the development of modifications.

A proposal has been written for the design and construction of a prototype. This vehicle would consist of the Chevrolet step van modified for

†This work supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199.

use by a disabled person in a motorized wheelchair. Driving control would be effected with an electro-hydraulic servo using a constant-pressure, variable-flow hydraulic system to operate steering, brake, and throttle mechanisms. Control signals to the servo systems will be processed by transfer functions set up on a small analog computer in the van.

The driver would operate the vehicle by manipulating finger-activated controls on a console while sitting in his wheelchair in the normal driving position. He would enter the vehicle by using a hydraulic lift and move to the driving console where a restraint mechanism will secure him and the wheelchair.

Studies in the design of control handles useable by disabled drivers, the proper configuration of the driver-restraint mechanism, and the determination of optimal control-system transfer functions would be part of the prototype vehicle development. Both linear and nonlinear transfer functions will be considered in order to solve the problem of obtaining large control dynamic range in the face of limited muscular capability.

17.3 Optical-Aural Technique for Reading by the Blind

The objective of this work is to develop a simple electronic aid which may enable sightless persons to read standard typewritten and printed material.

A light-sensitive pen and a straight edge (ruler) is proposed to manually scan the printed page line by line and produce sounds which are aesthetically acceptable and easily recognized as characteristic of individual letters in the alphabet. With training, it is hoped that the operator will be able to recognize sound groups as characteristic of words. The pen would have a straight column of uniformly-spaced sensing cells, from five to fifteen in number. The spacing and height of the column of cells would be manually adjustable to match the height of the typeface. Each cell would control a corresponding musical note appearing only when black typeface is encountered.

A "1" or an "I" would be recognized by all tones sounding simultaneously; a "V" by an arpeggio first descending, then ascending; an E by a burst of all tones followed by a "triad". Each letter would have its own aural characteristics. Allignment of the guide ruler and the printed line

would be achieved by correcting for any absence of highest or lowest notes at either end of an exploritory scan of the line.

Optical scanners with aural output have already been used by the blind with less than optimal success. However, these systems employed frequency variation to convey information, resulting in very unpleasant sounds which made it difficult for the user to concentrate. It is expected that the use of musical structures as cues to letter shape may minimize this problem.

17.4 References

1. Fariss, Gary B. and Burr Logeman, "A Highway Vehicle for Operation by a Quadriplegic," Report for GE 242, General Engineering Department, University of Illinois, Spring 1968.

H. Knoebel
W. D. Compton
V. Krone
J. Burt

18.1 Introduction

The ability to align and realign certain color centers with polarized light of one wavelength and the associated absorption or transmission of light of another wavelength for read out suggests their use in a memory device. Such a memory system would provide high storage density and high speed with which information may be imprinted and extracted.

18.2 M_A Color Centers in KCl

M_A color centers consist of two neighboring F centers pinned by a sodium impurity ion. This aggregate is table at liquid nitrogen temperatures and exists at high densities. They can be used as a binary memory [1] in the following way (see Fig. 18.1):

(1) Irradiate with 540 nm light polarized in the $\langle 110 \rangle$ direction. The initially randomly directed centers will realign so that 840 nm light polarized in the $\langle \bar{1}10 \rangle$ direction is not absorbed.

(2) Irradiate with 540 nm light polarized in the $\langle \bar{1}10 \rangle$ direction. The 840 nm light polarized in the $\langle 110 \rangle$ direction will now be absorbed. The 840 nm light does not reorientate the centers so it may be read without destruction of the memory. Fatigue due to cycling with 540 nm light does not become evident until many millions of cycles.

18.3 Acoustic Diffraction of Light

Bragg diffraction of light by a sound column in a solid or liquid medium is a promising means of steering a laser beam [2,3]. Deflection in x, y directions results from placing two Bragg deflectors at right angles to one another. In isotroic media, the diffraction angle is independent of the light polarization; hence, an electropic modulator can change the polarization of the initial laser beam allowing the writing and erasing to be done with the same apparatus. The reading system will require another deflection system.

---

†This work supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAB-07-67-C-0199.

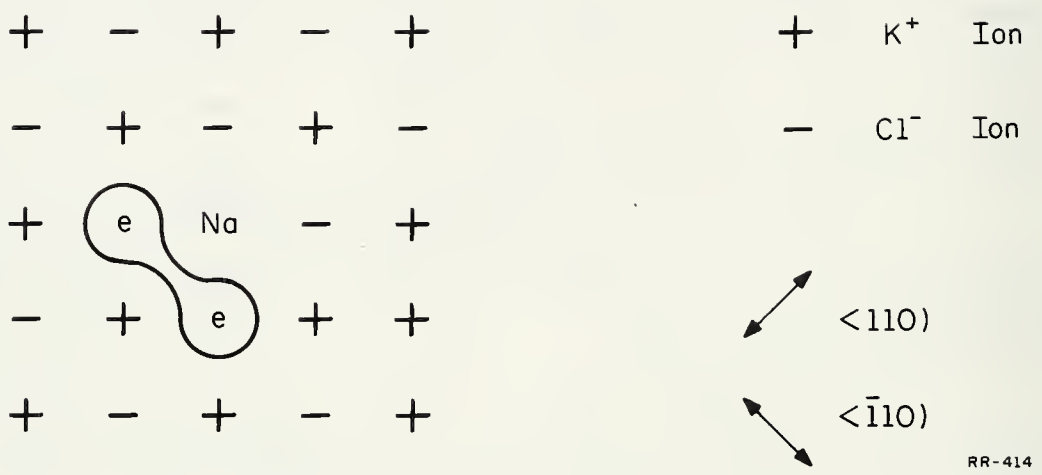


Fig. 18.1. M_a color center in KCl lattice.

In Fig. 18.2 a light beam of wavelength and diameter D is diffracted by the plane sound waves into an angle $\frac{1}{2}\theta_B = \lambda/\Lambda$ where Λ is the acoustic wavelength. The angular spread of the initial beam is λ/nD and the angular spread of the diffracted beam is almost $4\Lambda/L$ where n is the refractive index of the medium. The total number of resolvable spots is equal or less than $\tau\Delta f$ where τ is $(D/v)\csc\varphi$, v being the velocity of sound and Δf the bandwidth of the transducer. The intensity of the diffracted beam is proportional to $\sin^2\alpha$ and the intensity of the primary beam is $\cos^2\alpha$ where $\alpha = (\pi n^3 pL\epsilon)/2\lambda$, where p is the photoelastic constant, and ϵ is the maximum strain associated with the sound field.

Rapid access to large numbers of resolvable spots requires microwave frequencies which are unobtainable with conventional transducers other than thin films. However, certain anisotropic crystals such as litium niobate are piezoelectric and show low acoustic attention at these frequencies. The possibility of using such materials in a microwave cavity [4] or transmission line is being investigated as a means of generating the sound field. Since lithium niobate is also bifringent, light diffracted by it does not follow the ordinary Bragg laws [5]. Larger diffraction angles result from this.

Initially, a deflection system operating with conventional transducers will be build to evaluate the optical and acoustic problems present.

18.4 References

1. I. Schneider, Applied Optics, 6., 2197
2. E. Gordon and M. Cohen, Bell Sys. Tech. J., 44, 693
3. H. Hance, J. Parks, J. Acoustical Soc. America, 38, 14
4. De Klerk, Bolef, Trans. Sonics and Ultrasonics, 10, 27
5. R. Dixon, IEEE J. Quantum Electronics, 3, 85

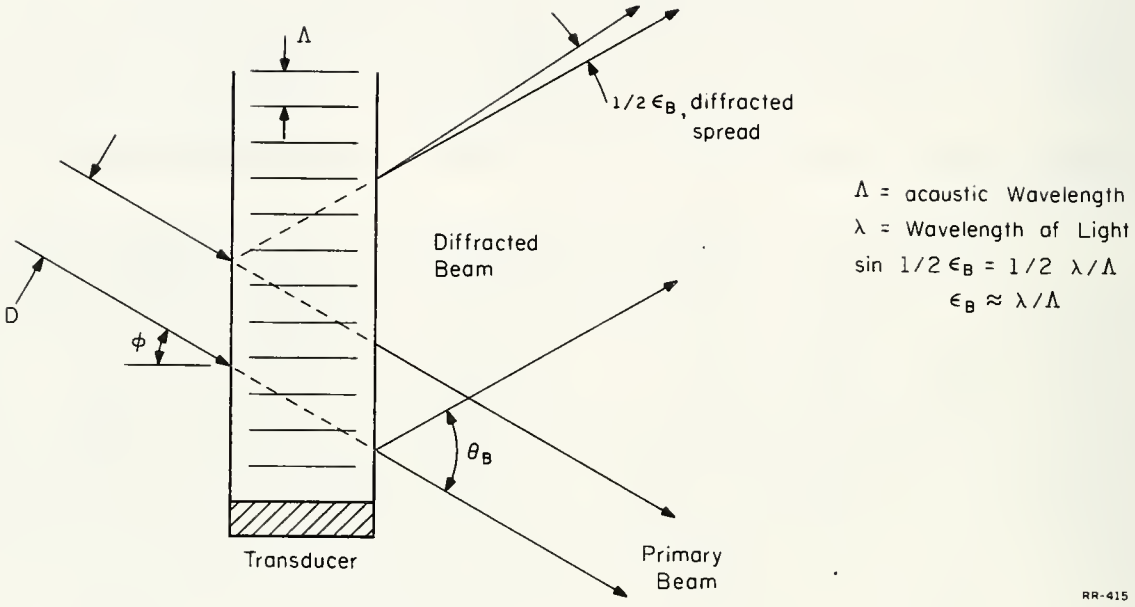


Fig. 18.2. Bragg diffraction cell.

Distribution List as of August 1, 1969

Dr A.A. Dougal
Asst Director (Research)
Ofc of Defense Res & Eng
Department of Defense
Washington, D.C. 20301

Office of Deputy Director
(Research and Information, Rm 3D1037)
Department of Defense
The Pentagon
Washington, D.C. 20301

Director, Advanced Research Projects
Agency
Department of Defense
Washington, D.C. 20301

Director for Materials Sciences
Advanced Research Projects Agency
Department of Defense
Washington, D.C. 20301

Headquarters
Defense Communications Agency (340)
Washington, D.C. 20305

Defense Documentation Center
Attn: DDC-TCA
Cameron Station
Alexandria, Virginia 22314 (50 Copies)

Director
National Security Agency
Attn: TDL
Fort George G. Meade, Maryland 20755

Weapons Systems Evaluation Group
Attn: Colonel Blaine O. Vogt
400 Army-Navy Drive
Arlington, Virginia 22202

Central Intelligence Agency
Attn: OCR/DD Publications
Washington, D.C. 20505

Hq USAF (AFRDD)
The Pentagon
Washington, D.C. 20330

Hq USAF (AFRDDG)
The Pentagon
Washington, D.C. 20330

Hq USAF (AFRDSD)
The Pentagon
Washington, D.C. 20330

Colonel E.P. Gaines, Jr.
ACDA/FO
1901 Pennsylvania Ave N.W.
Washington, D.C. 20451

Lt Col R.E. Kalisch (SREE)
Chief, Electronics Division
Directorate of Engineering Sciences
Air Force Office of Scientific Research
Arlington, Virginia 22209

Dr I.R. Mirman
AFSC (SCT)
Andrews Air Force Base, Maryland 20331

AFSC (SCTSE)
Andrews Air Force Base, Maryland 20331

Mr Morton M. Pavane, Chief
AFSC Scientific and Technical Liaison Office
26 Federal Plaza, Suite 1313
New York, New York 10007

Rome Air Development Center
Attn: Documents Library (EMTLD)
Griffiss Air Force Base, New York 13440

Mr H.E. Webb (EMMIIS)
Rome Air Development Center
Griffiss Air Force Base, New York 13440

Dr L.M. Hollingsworth
AFCRL (CRN)
L.G. Hanscom Field
Bedford, Massachusetts 01730

AFCRL (CRMPLR), Stop 29
AFCRL Research Library
L.G. Hanscom Field
Bedford, Massachusetts 01730

Hq ESD (ESTI)
L.G. Hanscom Field
Bedford, Massachusetts 01730 (2 copies)

Professor J. J. D'Azzo
Dept of Electrical Engineering
Air Force Institute of Technology
Wright-Patterson AFB, Ohio 45433

Dr H.V. Noble (CAVT)
Air Force Avionics Laboratory
Wright-Patterson AFB, Ohio 45433

Director
Air Force Avionics Laboratory
Wright-Patterson AFB, Ohio 45433

AFAL (AVTA/R.D. Larson
Wright-Patterson AFB, Ohio 45433

Director of Faculty Research
Department of the Air Force
U.S. Air Force Academy
Colorado Springs, Colorado 80840

Academy Library (DFSLB)
USAF Academy
Colorado Springs, Colorado 80840

Director
Aerospace Mechanics Division
Frank J. Seiler Research Laboratory (OAR)
USAF Academy
Colorado Springs Colorado 80840

Director, USAF PROJECT RAND
Via: Air Force Liaison Office
The RAND Corporation
Attn: Library D
1700 Main Street
Santa Monica, California 90045

Hq SAMSO (SMTTA/Lt Nelson)
AF Unit Post Office
Los Angeles, California 90045

Det 6, Hq OAR
Air Force Unit Post Office
Los Angeles, California 90045

AUL3T-9663
Maxwell AFB, Alabama 36112

AFETR Technical Library
(ETV,MU-135)
Patrick AFB, Florida 32925

AUTC (ADBPS-12)
Eglin AFB, Florida 32542

Mr E.R. Locke
Technical Adviser, Requirements
USAF Security Service
Kelly Air Force Base, Texas 78241

Hq AMD (AMR)
Brooks AFB, Texas 78235

USAFSAM (SMKDR)
Brooks AFB, Texas 78235

Commanding General
Attn: STEWS-RE-L, Technical Library
White Sands Missile Range
New Mexico 88002 (2 copies)

Hq AEDC (AETS)
Attn: Library/Documents
Arnold AFS, Tennessee 37389

European Office of Aerospace Research
APO New York 09667

Physical & Engineering Sciences Division
U.S. Army Research Office
3045 Columbia Pike
Arlington, Virginia 22204

Commanding General
U.S. Army Security Agency
Attn: IARD-T
Arlington Hall Station
Arlington, Virginia 22212

Commanding General
U.S. Army Materiel Command
Attn: AMCRD-TP
Washington, D.C. 20315

Technical Director (SMUFA-A2000-107-1)
Frankford Arsenal
Philadelphia, Pennsylvania 19137

Redstone Scientific Information Center
Attn: Chief, Document Section
U.S. Army Missile Command
Redstone Arsenal, Alabama 35809

Commanding General
U.S. Army Missile Command
Attn: AMSMI-REX
Redstone Arsenal, Alabama 35809

Commanding General
U.S. Army Strategic Communications Command
Attn: SCC-CG-SAE
Fort Huachuca, Arizona 85613

Commanding Officer
Army Materials and Mechanics Res. Center
Attn: Dr H. Priest
Watertown Arsenal
Watertown, Massachusetts 02172

Commandant
U.S. Army Air Defense School
Attn: Missile Science Division, C&S Dept
P.O. Box 9390
Fort Bliss, Texas 79916

Commandant
U.S. Army Command & General Staff College
Attn: Acquisitions, Library Division
Fort Leavenworth, Kansas 66027

Commanding Officer
U.S. Army Electronics R&D Activity
White Sands Missile Range, New Mexico 88002

Mr Norman J. Field, AMSEL-RD-S
Chief, Office of Science & Technology
Research and Development Directorate
U.S. Army Electronics Command
Fort Monmouth, New Jersey 07703

Commanding Officer
Harry Diamond Laboratories
Attn: Dr Berthold Altman (AMXDO-TI)
Connecticut Avenue and Van Ness St N.W.
Washington, D.C. 20438

Director
Walter Reed Army Institute of Research
Walter Reed Army Medical Center
Washington, D.C. 20012

Commanding Officer (AMXRD-BAT)
U.S. Army Ballistics Research Laboratory
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Technical Director
U.S. Army Limited War Laboratory
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Commanding Officer
Human Engineering Laboratories
Aberdeen Proving Ground
Aberdeen, Maryland 21005

U.S. Army Munitions Command
Attn: Science & Technology Br. Bldg 59
Picatinny Arsenal, SMUPA-VA6
Dover, New Jersey 07801

U.S. Army Mobility Equipment Research
and Development Center
Attn: Technical Document Center, Bldg 315
Fort Belvoir, Virginia 22060

Director
U.S. Army Engineer Geodesy,
Intelligence & Mapping
Research and Development Agency
Fort Belvoir, Virginia 22060

Dr Herman Robl
Deputy Chief Scientist
U.S. Army Research Office (Durham)
Box CM, Duke Station
Durham, North Carolina 27706

Richard O. Ulsh (CRDARD-IPO)
U.S. Army Research Office (Durham)
Box CM, Duke Station
Durham, North Carolina 27706

Mr Robert O. Parker, AMSEL-RD-S
Executive Secretary, JSTAC
U.S. Army Electronics Command
Fort Monmouth, New Jersey 07703

Commanding General
U.S. Army Electronics Command
Fort Monmouth, New Jersey 07703
Attention:AMSEL-SC
RD-GF
RD-MT
XL-D
XL-E
XL-C
XL-S (Dr R. Buser)
HL-CT-DD
HL-CT-R
NL-CT-L (Dr W.S. McAfee)
HL-CT-O
HL-CT-I
NL-CT-A
NL-D
NL-A
NL-P
NL-P-2 (Mr D. Harsts)
NL-R (Mr R. Kulinyi)
NL-S
KL-D
KL-E
KL-S (Dr H. Jacobs)
KL-SM (Drs Schiel/Hieslmair)
KL-T
VL-D
VL-F (Mr R.J. Niemela)
WL-D
(1 copy to each symbol listed individually addressed)

Dr A.D. Schnitzler, AMSEL-HL-NVII
Night Vision Laboratory, USAECOM
Fort Belvoir, Virginia 22060

Dr G.M. Janney, AMSEL-HL-NVOR
Night Vison Laboratory,USAECOM
Fort Belvoir, Virginia 22060

Atmospheric Sciences Office
Atmospheric Sciences Laboratory
White Sands Missile Range
New Mexico 88002

Missile Electronic Warfare,
Technical Area, AMSEL-WT-MT
White Sands Missile Range
New Mexico 88002

Project Manager
Comman Positioning & Navigation Systems
Attn: Harold H. Bahr (AMCPM-NS-TM), Bldg 439
U.S. Army Electronics Command
Fort Monmouth, New Jersey 07703

Director, Electronic Programs
Attn: Code 427
Department of the Navy
Washington, D.C. 20360

Commander
U.S. Naval Security Group Command
Attn: G43
3801 Nebraska Avenue
Washington, D.C. 20390

Director
Naval Research Laboratory
Washington, D.C. 20390
Attn: Code 2027 6 copies
Dr W.C. Hall, Code 7000 1 copy
Dr A. Brodzinsky, Sup.Elec Div. 1 copy

Dr C.M.R. Winkler
Director, Time Service Division
U.S. Naval Observatory
Washington, D.C. 20390

Naval Air Systems Command
AIR 03
Washington, D.C. 20360 2 copies

Naval Ship Systems Command
Ship 031
Washington, D.C. 20360

Naval ship Systems Command
Ship 035
Washington, D.C. 20360

U.S. Naval Weapons Laboratory
Dahlgren, Virginia 22448

Naval Electronic Systems Command
ELEX 03, Room 2046 Munitions Building
Department of the Navy
Washington, D.C. 20360 (2 copies)

Commander
Naval Electronics Laboratory Center
Attn: Library
San Diego, California 92152 (2 copies)

Deputy Director and Chief Scientist
Office of Naval Research Branch Office
1030 Eat Gree Street
Pasadena, California 91101

Library (Code2124)
Technical Report Section
Naval Postgraduate School
Monterey, California 93940

Glen A. Myers (Code 52Mv)
Assoc Professor of Elec. Engineering
Naval Postgraduate School
Monterey, California 93940

Commanding Officer and Director
U.S. Naval Underwater Sound Laboratory
Fort Trumbull
New London, Connecticut 06840

Commanding Officer
Naval Avionics Facility
Indianapolis, Indiana 46241

Dr H. Harrison, Code RRE
Chief, Electrophysics Branch
National Aeronautics & Space Admin.
Washington, D.C. 20546

NASA Lewis Research Center
Attn: Library
21000 Brookpark Road
Cleveland, Ohio 44135

Los Alamos Scientific Laboratory
Attn: Report Library
P.O. Box 1663
Los Alamos, New Mexico 87544

Federal Aviation Administration
Attn: Admin Stds Div (MS-110)
800 Independence Ave S.W.
Washington, D.C. 20590

Head, Technical Services Division
Naval Investigative Service Headquarters
4420 North Fairfax Drive
Arlington, Virginia 22203

Commander
U.S. Naval Ordnance Laboratory
Attn: Librarian
White Oak, Maryland 21502 (2 copies)

Commanding Officer
Office of Naval Research Branch Office
Box 39 FPO
New York, New York 09510

Commanding Officer
Office of Naval Research Branch Office
219 South Dearborn Street
Chicago, Illinois 60604

Commanding Officer
Office of Naval Research Branch Office
495 Summer Street
Boston, Massachusetts 02210

Commander (ADL)
Naval Air Development Center
Johnsville, Warminster, Pa 18974

Commanding Officer
Naval Training Device Center
Orlando, Florida 32813

Commander (Code 753)
Naval Weapons Center
Attn: Technical Library
China Lake, California 93555

Commanding Officer
Naval Weapons Center
Corona Laboratories
Attn: Library
Corona, California 91720

Commander, U.S. Naval Missile Center
Point Mugu, California 93041

W.A. Eberspacher, Associate Head
Systems Integration Division
Code 5340A, Box 15
U.S. Naval Missile Center
Point Mugu, California 93041

Mr M. Zane Thornton, Chief
Network Engineering, Communications
and Operations Branch
Lister Hill National Center for
Biomedical Communications
8600 Rockville Pike
Bethesda, Maryland 20014

U.S. Post Office Department
Library - Room 1012
12th & Pennsylvania Ave, N.W.
Washington, D.C. 20260

Director
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Mr Jerome Fox, Research Coordinator
Polytechnic Institute of Brooklyn
55 Johnson Street
Brooklyn, New York 11201

Director
Columbia Radiation Laboratory
Columbia University
538 West 120th Street
New York, New York 10027

Director
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Director
Stanford Electronics Laboratories
Stanford University
Stanford, California 94305

Director
Microwave Physics Laboratory
Stanford University
Stanford, California 94305

Director, Electronics Research Laboratory
University of California
Berkeley, California 94720

Director
Electronic Sciences Laboratory
University of Southern California
Los Angeles, California 90007

Director
Electronics Research Center
The University of Texas at Austin
Austin Texas 78712

Division of Engineering and Applied Physics
210 Pierce Hall
Harvard University
Cambridge, Massachusetts 02138

Dr G.J. Murphy
The Technological Institute
Northwestern University
Evanston, Illinois 60201

Dr John C. Hancock, Head
School of Electrical Engineering
Purdue University
Lafayette, Indiana 47907

Dept of Electrical Engineering
Texas Technological College
Lubbock, Texas 79409

Aerospace Corporation
P.O. Box 95085
Los Angeles, California 90045
Attn: Library Acquisitions Group

Proffessor Nicholas George
California Inst of Technology
Pasadena, California 91109

Aeronautics Library
Graduat Aeronautical Laboratories
California Institute of Technology
1201 E. California Blvd
Pasadena, California 91109

The John Hopkins University
Applied Physics Laboratory
Attn: Document Librarian
8621 Georgia Avenue
Silver Spring, Maryland 20910

Raytheon Company
Attn: Librarian
Bedford, Massachusetts 01730

Raytheon Company
Research Division Library
28 Seyon Street
Waltham, Massachusetts 02154

Dr Sheldon J. Wells
Electronic Properties Information Center
Mail Station E-175
Hughes Aircraft Company
Culver City, California 90230

Dr Robert E. Fontana
Systems Research Laboratories Inc.
7001 Indian Ripple Road
Dayton, Ohio 45440

Nuclear Instrumentation Group
Bldg 29, Room 101
Lawrence Radiation Laboratory
University of California
Berkeley, California 94720

Sylvania Electronic Systems
Applied Research Laboratory
Attn: Documents Librarian
40 Sylvan Road
Waltham, Massachusetts 02154

Hollander Associates
P.O. Box 2276
Fullerton, California 92633

Illinois Institute of Technology
Dept of Electrical Engineering
Chicago, Illinois 60616

The University of Arizona
Dept of Electrical Engineering
Tucson, Arizona 85721

Utah State University
Dept Of Electrical Engineering
Logan, Utah 84321

Case Institute of Techoology
Engineering Division
University Circle
Cleveland, Ohio 44106

Hunt Library
Carnegie-Mellon University
Schenley Park
Pittsburgh, Pennsylvania 15213

Dr Leo Young
Stanford Research Institute
Menlo Park, California 94025

School of Engineering Sciences
Arizona State University
Tempe, Arizona 85281

Engineering & Mathematical Sciences Library
University of California at Los Angeles
405 Hilgred Avenue
Los Angeles, California 90024

The Library
Government Publications Section
University of California
Santa Barbara, California 93106

Carnegie Institute of Technology
Electrical Engineering Department
Pittsburgh, Pennsylvania 15213

Professor Joseph E. Rowe
Chairman, Dept of Electrical Engineering
The University of Michigan
Ann Arbor, Michigan 48104

New York University
College of Engineering
New York, New York 10019

Syracuse University
Dept of Electrical Engineering
Syracuse, New York 13210

Yale University
Engineering Department
New Haven, Connecticut 06520

Airborne Instruments Laboratory
Deerpark, New York 11729

Raytheon Company
Attn: Librarian
Bedford, Massachusetts 01730

Lincoln Laboratory
Massachusetts Institute of Technology
Lexington, Massachusetts 02173

The University of Iowa
The University Libraries
Iowa City, Iowa 52240

Lenkurt Electric Co,Inc
1105 County Road
San Carlos, California 94070
Attn: Mr E.K. Peterson

Philco Ford Corporation
Communications & Electronics Div.
Union Meeting and Jolly Rods
Blue Bell, Pennsylvania 19422

Union Carbide Corporation
Electronic Division
P.O. Box 1209
Mountain View, California 94041

Electromagnetic Compatibility Analysis Center
(ECAC), Attn: ACLP
North Severn
Annapolis, Maryland 21402

Director
U. S. Army Advanced Materiel Concepts Agency
Washington, D.C. 20315

ADDENDUM

Dept of Electrical Engineering
Rice University
Houston, Texas 77001

Research Laboratories for the Eng. Sc.
School of Engineering & Applied Science
University of Virginia
Charlottesville, Virginia 22903

Dept of Electrical Engineering
College of Engineering & Technology
Ohio University
Athens, Ohio 45701

Project Mac
Document Room
Massachusetts Institute of Technology
545 Technology Square
Cambridge, Massachusetts 02139

Lehigh University
Dept of Electrical Engineering
Bethelem, Pennsylvania 18015

Commander Test Command (TCDT-)
Defense Atomic Support Agency
Sandia Base
Albuquerque, New Mexico 87115

Materials Center Reading Room 13-2137
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Professor James A. Cadzow
Department of Electrical Engineering
State University of New York at Buffalo
Buffalo, New York 14214

ERRATUM

Mr Jerome Fox, Research Coordinator
Polytechnic Institute of Brroklyn
55 Johnson St (Should be 333 Jay St)
Brooklyn, N.Y. 11201

OMIT

Mr Morton M. Pavane, Chief
AFSC Scientific & Tech. Liaison Office
26 Federal Plaza, Suite 1313
New York, New York 10007

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
University of Illinois Coordinated Science Laboratory Urbana, Illinois 61801	Unclassified
	2b. GROUP

3. REPORT TITLE

PROGRESS REPORT FOR SEPTEMBER 1968 THROUGH JUNE 1969

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1, 1969	352	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
DAAB-07-67-C-0199 b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Joint Services Electronics Program thru U. S. Army Electronics Command Fort Monmouth, New Jersey 07703

13 ABSTRACT

During the past year, three experiments have been pursued to the stage of the preparation of Ph. D. theses. These are electron-impact desorption of ions and neutrals, energy-angular distribution of secondary electrons, and high-resolution electron spectroscopy of solid surfaces. The thesis for the first of these has been completed and has been presented as a CSL report. The second two are in preparation. Brief summaries of the work covered in these theses are given in the present progress report. A new activity in surface research has been started in the laboratory under the direction of Professor Gert Ehrlich. This work emphasizes the study of the chemistry of the gas-solid interface. A brief description of the types of experiments which will be pursued is given.

Resistivity and mobility data as a function of temperature, and photo-luminescent spectra are given foe vacuum-deposited CdS films about one micron thick. Measurements on metal-insulator-CdS structures show light emission at voltages as low as one volt, making a tunnel injection mechanism doubtful. Parameters are discussed for experimental measurements of barries heights in tunneling experiments. The behavior of an omegatron total-partial pressure gauge is discussed, as well as the effect of small magnetic fields on typical ionization gauges. The scanning electron microscope has been used to examine surfaces prepared by electrical-discharge machining. The results have led to cutting conditions improving the rate, and reducing surface damage.

(Cont'd on next 2 pages)

DD FORM 1 NOV 65 1473

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Surface Physics						
Applied Physics						
Plasma Physics						
Rarefied-Gas Dynamics						
High-Voltage Breakdown						
Space and Atmospheric Sciences						
Semiconductor Physics						
Computer						
Plasma Display						
Urban Development						
Information Science						
Networks						
Control Systems						
Switching Systems						
Cognitive Memory						
Seismic Vehicle Detection						
Engineering Aids for the Handicapped						
Photo-Chromic Memory for Computer Applications						

The effect of collisions on the instability of cold and warm electron beams in a plasma has been investigated experimentally. It is found that the instability is effectively suppressed for velocity spreads much smaller than previously assumed. Computer calculations are presented that support this finding. The spatial-temporal development of the instability in the presence of boundaries is discussed theoretically.

More accurate Monte-Carlo solutions of the Boltzmann equation have been obtained for the pseudoshock and the heat-transfer problems by using the new techniques developed. These solutions and the solution of the shock wave problem enable us to make detailed studies of the nonequilibrium gas behavior in rarefied-gas flow problems. These studies include the investigation of the effect of the gas-surface interaction on the gas behavior near a solid surface. Because of the improvement of the accuracy of evaluating the Boltzmann collision integrals, we are able to extend our study to near-equilibrium conditions. The second-order effect of the Chapman-Enskog's distribution function on the collision integral is being studied. We have conducted more error studies of our Monte-Carlo method to establish a firm basis for tackling more complex problems. We have completed the initial phase of our plans to extend the application of our Monte-Carlo method to more complex problems by using a large-scale parallel-operation computer.

The nature of the cathode protrusions responsible for initiation of electrical breakdown in ultrahigh vacuum has been studied by means of the scanning electron microscope. These protusions appear to be fingers of metal extending from the edges or interiors of small craters. The craters are produced when clumps of charged material, as yet unidentified, impinge on the surface of the cathode (or the anode), melting the local region. Using drastic cleaning procedures, the clumps, under laboratory conditions, appear to have been eliminated, thus eliminating the protusions and permitting the electrode voltage to be raised one to two orders of magnitude higher, without field emission or electrical breakdown, than was previously practical. In several cases, 60 kV has been sustained between tungsten electrodes spaced 0.04mm apart, indicating a field of about 15×10^8 V/m at the electrode surface.

A statement of the status of the feasibility study for the orgiting-gyro relativity satellite is given. The program and objectives of the study of another orbiting-gyro experiment designed to measure gas density and gas-surface interaction parameters are given. The results of a preliminary study of a composite-shaped satellite are given which illustrate the feasibility of the proposed atmospheric experiment. The resulting equations of a more complete aerodynamic analysis of spinning disks, cones, and cylinders at arbitrary angles of attack and the evaluation of these results using numerical techniques are discussed. The fesibility of measuring the rates of adsorption and desorption or neutral molecules on the satellite surface is evaluated as a possible secondary objective in the proposed atmospheric experiment. Molecular vibrations of cyclododeca-sulfur, S_{12}, were interferometrically measured in the far infrared using a polarized source.

Defects and impurities determine many of the properties of semiconductors. In many cases, radiation offers a convenient means of introducing defects. Results are presented of the influence or irradiation upon the recombination luminescence of silicon, the luminescent spectra of CdSe when excited into laser action, and the electrical properties of p-i-n diodes. The tunneling between metal-semiconductor junctions is investigated as a function of the impurity level in the semiconductor.

Arithmetic and code-checking routines have been added to the ILLAR system. The FORTRAN compiler has been modified to make subprograms recursive (i.e.,they can call themselves). The CSLx programming package has been completed, and the Mitchell 16-mm movie production facility has been installed. The display system has been modified to provide 256 intensity levels, and a programming system that uses this feature has been developed to produce prospective drawings in which hidden planes are eliminated. Examples are shown

The appearance of a 128 X 128 Plasma Display Panel connected to a digital computer is shown. Studies of two-on-state operation, applications to memory, circuits, field calculations, stability, and color in the Plasma Display Panel are discussed.

An in-depth study of the factors causing the growth and decay of Kankakee, Illinois has been started. During this first phase of the study, data have been collected and prepared for computer display on the location, price, and date of real estate transactions throughout the Kankakee metropolitan area. Plans for the analysis of these data are described.

Progress in the area of information-processing techniques and systems is reported. Topics discussed include methods of achieving systems reliability through coding, design techniques of digital systems, estimation and filtering, source coding and time-continuous channels, information retrieval and connunications.

A computer optimization method is reported for the broad-band matching of complex generator and loat impedances toa lumped-distributed network. The network may contain active elements such as transistors. The study of system diagnosis from the topological point of view is in progress. In addition some results from the study of the characterization of biplanar graphs are given.

Intensive investigations of computationally feasible methods of design of large-scale systems is underway. Furthermore, additional results concerning minimax algorithms and feedback straegies for differential games have been obtained. Finally, suboptimal filtering, sensitivity of distributed parameter systems, adaptive systems, and applications of optimal control theory to traffic control and rolling-mill design are being investigated.

The investigation of fault diagnosis methods for digital systems has been continued. Results have been obtained on the improvement of diagnosability by test-point insertion, on the detection of multiple faults by single fault tests, and on the generation of diagnostic tests for networks of modules. Work continues on methods of synthesizing sequential networks of modules. Work continues on methods of synthesizing sequential networks which are optimal from the diagnosis viewpoint and on fault-diagnostic properties of regular arrays.

Fundamental properties of perceptive cognition are investigated. Semantic techniques for parsing and compliling are developed. Architecture of an associative processor is presented. Other progress relating to cognitive structures are also reported.

A telemetered angine clock signal is being used in an attempt to determine if vehicular seismic signals contain well-defined frequency components related to engine speed. Preliminary results of an attempt to determine the weight of passing vehicles from an analysis of their seismic signals is described.

A preliminary design is described for a highway vehicle to be operated by a severely disabled person. A vehicle is being donated, and a proposal has been written for further study and the modification of the donated vehicle. A technique which may enable blind to read newsprint is described.

A proposed mechanism for using M_A color centers in a write, erase, and nondestructive read-out memory system is described.